Developing and Articulating a Process to Build and Validate a Measurement Rubric for Measuring Creativity in Physics in High Schools

Weiping Xu

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

Centre for Educational Measurement and Assessment

Sydney School of Education and Social work

Faculty of Arts and Social Sciences

The University of Sydney

2024

Statement of originality

This is to certify that, to the best of my knowledge, the content of this thesis is my own work.

This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Weiping Xu

March 2024

Acknowledgements

"Suddenly turning around, the one is already in the dimly lit place of lanterns." —Qiji Xin

Inspired by this poignant verse, I begin to reflect deeply on my gratitude to those who have illuminated the path of my academic pursuits.

My deepest appreciation goes to my esteemed supervisor, Professor Jim Tognolini, who has been like a guiding light in the lantern-lit dusk, providing wisdom and support throughout the intricacies of my research journey. I also thank my auxiliary supervisor, Dr Kathryn Bartimote, for developing my independent thinking and problem-solving skills to enable me to become an independent researcher. I express my deep gratitude to Dr. Joh S. Twing and Associated Professor Sofia Kesidou for their professional review and invaluable contributions that led to the overall improvement of the thesis.

The hallowed halls of the University of Sydney have always been an academic mecca where the lanterns of knowledge light the way forward. I am grateful to this institution and the Sydney School of Education and Social Work for providing me with an environment conducive to intellectual growth.

To my colleagues and fellow travellers on this academic road, especially members at the Centre for Educational Measurement and Assessment – Kathleen Walker, Shafiza binti Mohamed, Rayanne Shakra, Kuliga Tanasavate, Abdullah Alzhrany, Jay Lim, Li Liang and Tzu-ping Lu – your shared insights and camaraderie have made the journey so much more fulfilling and meaningful, akin to the shared warmth in the lantern-lit spaces.

I would like to dedicate this thesis to my loving family, who have always supported

me with their unconditional love and encouragement. They have taught me the value of perseverance, resilience and honesty, and inspired me to keep pursuing my passion for research.

I am deeply grateful for the love and support my boyfriend has shown me. We faced the storms and chaos together and never lost our courage and hope.

Thanks to Kerry Davies AE for professionally copyediting this thesis in accordance with the guidelines of the university and the Institute of Professional Editors, through which Kerry is an Accredited Editor (AE). While every effort has been made to correct inadvertent errors or apparent contradictions in this thesis, I acknowledge that any errors that remain are solely my own.

Finally, I would like to express my gratitude to myself. Reflecting on the preceding four years, I undertook a solitary transoceanic expedition, traversing vast distances to arrive at this unfamiliar nation and city. The adversities encountered during this journey defy facile articulation to outsiders. Notwithstanding the multitude of challenges, on retrospective contemplation it becomes apparent that each trial has, in essence, forged the contours of the individual I embody at present.

Abstract

High school teachers with some knowledge of creativity often have difficulty identifying the creativity of their students, although there are several psychometric instruments available to assess creativity. As a result, teachers are unable to assess the effectiveness of their teaching in terms of creativity, and students' creativity is not targeted and enhanced in their teaching. This study aims to develop a measurement rubric for creativity that describes, through performance standards, what students should know and can do in relation to creativity along a development continuum. A measurement rubric is construct-oriented and shows growth or progress in what is being measured over time. A measurement rubric for creativity allows teachers to design assessment tasks that are appropriate to their teaching practice, and one where student responses can be used as evidence to locate students on a developmental continuum of creativity by referencing them to the predetermined performance standards. Hence, these predetermined standards operationalise the definition of the construct of creativity.

This study follows the procedures proposed by Tognolini (2018) for constructing a measurement rubric to define the construct of creativity and its growth in students. This procedure includes defining the construct to be measured, determining the components of the construct and developing performance standards for each level of performance on each component. The definition of creativity, its components and the associated performance standards establish a basis for a draft generic measurement rubric for creativity that might be useful across disciplines. Validity evidence for this draft generic measurement rubric for creativity was then collected for use in various disciplines (e.g., science and physics)

primarily through three focus groups and two teacher development workshops mainly focusing on content and construct evidence. Qualitative content analysis of this commentary provided validity evidence for the use of the draft generic measurement rubric for creativity. Based on the results of the analysis, the draft generic measurement rubric for creativity was amended and finalised to make it more valid as parts of the construction process of the rubric. Assessment tasks were then designed to generate measures applicable to the measurement rubric. These assessment tasks were examined to see what evidence they generated for the efficacy of the measurement rubric, which is another key evidence component for both content and construct validity. Assessment tasks were constructed through two teacher development workshops involving high school physics teachers in China, where they designed appropriate tasks for each component of creativity in the generic measurement rubric that could be used in the classroom to identify and locate students' response along the continuum described by the rubric.

The study found that participants in the focus group discussions and the teacher development workshops consensually agreed on the use of a generic measurement rubric for creativity in classroom assessment. They considered that this rubric is more consistent with the nature of creativity (e.g., thinking out of the box) and allowed for the design of a variety of assessment activities that incorporate interdisciplinary or transdisciplinary learning to promote the development of student creativity. The process of constructing a measurement rubric for creativity and its validation evidence provides new opportunities to focus on measuring student growth of creativity, and will likely generalise to other "21st century" constructs.

Contents

List of Tables

Glossary of terms

construct: Measurable and learnable capability of an individual.

component: constituent attributes to be measured that align with the definition of the construct being measured.

criterion referencing: Process of giving meanings to a student's performance by referencing their achievements to task-specific or course-specific criteria. See also *norm referencing*, *standard referencing*.

domain: Any specific field of knowledge. In education, curricula are developed for disciplines or subject domains, such as history, science, or mathematics.

domain generality: Universal characteristics or capacities independent of disciplinary or subject domain features. In schools, domain-generic ability can be taught and learned in various subject classes.

domain specificity: Specific characteristics or capacities strongly relevant to a discipline's feature.

domain-specific measurement rubric: Measurement scale used to locate students' performance based on how much they can do within a specific domain, or discipline. See also *generic measurement rubric*, *measurement rubric*.

generic measurement rubric: Measurement scale that can be used to locate students' performance, regardless of discipline or subject domain. See also *domain-specific measurement rubric*, *measurement rubric*.

measurement rubric: Scale for measuring all types of performance based on "how much" a student performs on a predetermined construct as captured in the explication of the

performance standards used to define growth in relation to the construct being measured. In this thesis, it is also referred to as a rubric. See also *domain-specific measurement rubric*, *generic measurement rubric*.

norm referencing: Process of giving meaning to a student's performance by referencing it to the performance of a normative group of peers. See also *criterion referencing*, *standard referencing*.

performance-level descriptor: Description of the learning outcome that students can achieve at each level of performance.

performance standard: Sequence of performance descriptors that show incremental growth along the developmental continuum for a specific construct.

standard referencing: Process of giving meanings to a student's performance by referencing it to a set of predetermined standards of performance. See also *criterion referencing, norm referencing*.

Chapter 1 Introduction

1.1 Creativity

Creativity has been described as one of the most important competencies of the 21st century because it enables people to adapt to a challenging future and unforeseen changes (Ananiadoui & Claro, 2009; Lamri & Lubart, 2021; Lucas, 2019). The National Education Association of the United States (2009) has identified creativity as one of four essential skills (along with critical thinking, communication and collaboration), also referred to as the "4Cs", that need to be supported by a broader educational base. This includes the development of new standards and assessments, as well as improvements in the curriculum and pedagogy. Over recent years a significant amount of work has been carried out to identify the cognitive and affective mechanisms underlying creativity (Beghetto & Kaufman, 2014), which has stretched to suggesting strategies to nurture creativity in classroom teaching (Beghetto, 2013; Beghetto & Kaufman, 2014; Blamires & Peterson, 2014; Cremin et al., 2006; Nuhoglu & Akgül, 2019; Sun et al., 2019b, 2020).

In addition, creativity has been incorporated into several government policies and associated curricula internationally to promote creativity education (Care, Vista, et al., 2019; Craft, 2006; Lucas, 2022; Patston et al., 2021). In a study of national education documents on "21st-century" skills across the world (UNESCO, 2016), the term "creativity" was mentioned in more than 50 countries. For example, in England, creativity was mentioned as one of the aims in the National Curriculum (Department for Education, 2014). More detailed descriptions of creativity can be found in the UK Qualifications and Curriculum Authority's

documentation, in which they advocate for the integration of creativity education into all national curricula and provide a basis for developing a fundamental knowledge of creativity in education (QCA, 2005). In addition, the framework of Personal, Learning and Thinking Skills (PLTS) identifies the focus and characteristics of a creative thinker (QCA, 2009). In Australia, the Australian Curriculum, Assessment and Reporting Authority (ACARA) places a significant emphasis on nurturing the fundamental cognitive aspect of creativity – creative thinking – and underscores it in conjunction with critical thinking as an important general capability in the Australian Curriculum (ACARA, 2016b). It also provides an explicit definition and an associated learning continuum for the combined construct of critical and creative thinking. The Organisation for Economic Cooperation and Development (OECD) Centre for Educational Research and Innovation (CERI) has also recognised the important role of creativity and has undertaken research to develop an assessment framework for the inclusion of creativity assessments for 15-year-old students in the Programme for International Student Assessment (PISA) (Lucas et al., 2013; OECD, 2019; Vincent-Lancrin, 2015). The PISA 2022 has incroporated and implemented an assessment of creative thinking was (Vincent-Lancrin et al., 2019).

The desire to foster creativity is also driving the direction of curriculum reform in China. Creativity is considered one of the core literacies of Chinese student development (Core Literacy Research Group, 2016). The Curriculum Standards for High Schools subsequently put forward the requirement of cultivating students' creativity as part of scientific literacies and competencies for lifelong learning (Ministry of Education of People's Republic of China, 2017a). In alignment with the Curriculum Standards for the High School

program, the Standard of the Physics Curriculum for Senior High Schools includes creativity and questioning as essential components of scientific thinking, and provides guidance that creativity and questioning should be embedded in science learning more generally, for example with scientific research (such as scientific discoveries and modelling), technology development, experimental design and improvement, and problem solving (Ministry of Education of People's Republic of China, 2017b).

Several empirical studies indicate that student growth in creativity measured by assessing and monitoring student performance remains a challenge in teaching practice (Andiliou & Murphy, 2010; Bereczki & Karpati, 2018; Kettler et al., 2018). During interviews with 20 physics teachers in China, Xu (2018) discovered that identifying and monitoring students' development in creativity during routine physics teaching proved to be a challenging task for teachers when designing assessment activities. Tran and his colleague (Tran et al., 2016) found that secondary school teachers in Vietnam lacked the basic understanding of creativity and the skills of teaching and assessing creativity, having conducted 234 teacher surveys and several in-depth interviews, and examined the lesson plans of the participants in the study. Similarly, Harris and Bruin (2017) suggested that school teachers in Australia still lack the ability to develop, recognise and assess students' creative and critical thinking as required by the national curriculum. An updated survey of a large sample of 631 teachers in Australia suggested significant increases in their knowledge of creativity; however, there was still a gap between the teachers' understanding of creativity and the definition provided in the literature (Cropley et al., 2019).

Given the urgent need for effective and applicable measures of creativity in

classrooms by teachers, the purpose of this thesis is to develop a measurement rubric provide evidence of student performance in relation to creativity, which enables teachers to locate students along a developmental continuum of what students know and can do in relation to creativity. Using measurement rubrics has an advantage that they can link assessment with effective pedagogy and thus enhance teachers' understanding of student creativity and at the same time provide them with the ability to monitor the impact of their teaching on student performance. Once this occurs, it should be possible to cultivate, teach and evaluate student performance in creativity as readily as any other construct in everyday teaching.

1.2 Assessment in Creativity

The "Four Ps" creativity framework – person, press (environment), product and process – conceptualise and investigate creativity by capturing its four significant influences (Rhodes, 1961). The "Four Ps" framework is extensively used to summarise and critically review different theories of creativity and its measurements (Bapna et al., 2017; Barbot, 2019; Said Metwaly et al., 2017). The assessment of "press", which observes environmental impact on creativity, can help teachers create an open classroom that allows creativity to blossom. However, it cannot provide evidence to demonstrate what students know and can do in relation to creativity. Hence, in the sections that follow, "press" has been set aside and existing measures of creativity are reviewed in relation to the three remaining categories of creativity: person, product, and process.

1.2.1 Person

Person refers to an individual's personality, motivation, attitudes, intellect, and other personal influences on creativity. The extent to which a person is considered creative has traditionally been measured by self-reporting instruments using either a yes-or-no format or on a 5-point Likert rating scale. Three of these measures are the Group Inventory for Finding Interests (GIFFI) (Davis & Rimm, 1982); the Five-Factor Personality Inventory (Hendriks et al., 1999); and the Eysenck Personality Scale (EPQ) (Eysenck & Eysenck, 1975). Some typical personality traits that exist across domains and cultures have been integrated into these measures and models. For example, the Big Five Model (BFM) (Goldberg, 2013; John et al., 2008) includes separate traits of neuroticism, extraversion, agreeableness, conscientiousness and openness in relation to personality. The main requirement of the BFM, that personality comprises five traits and these are entirely independent of each other, has been challenged by results of statistical analyses (DeYoung, 2007; DeYoung et al., 2002). As a result, alternative models have evolved that aggregate or merge interrelated traits – for example, openness and extraversion have been merged to produce a dimension of plasticity; and neuroticism, agreeableness and consciousness have been merged to produce the dimension of stability (Feist, 2019). This process of reconceptualising personality traits into "thicker dimensions" has produced the Big Two Model (BTM).

Instead of portraying a clear set of traits, creative individuals tend to contradict themselves because they better adjust to different situations and task purposes. Csikszentmihalyi (1996) attributed several paradoxical personality traits to creative celebrity, such as being: energetic and quiet; smart and naive; playful and responsible; imaginative and

realistic; introverted and extraverted; humble and proud. Moreover, the contrasting results of related studies reflect the ambiguous relationship between "person" and creativity (Feist, 2010). On one hand, the relationship between personality indicators in the BFM and creativity is supported by creativity research on employees (Yao & Li, 2021), and college students (Angulo-Brunet & Soto-Sanfiel, 2022). These results reinforce the idea that personality traits are among the reliable predictors of creative achievement (Batey & Furnham, 2006; Feist & Barron, 2003). However, this claim has also been refuted by other empirical studies. Findings from a correlation analysis between the BFM and creativity as measured by the Test for Creative Thinking-Drawing Production (TCT-DP) for college students indicate that none of the personality traits, except for the openness trait, consistently correlates with test results for creativity (Dollinger et al., 2004). One study distinguishes between this correlation in the arts and the correlation evidenced in the sciences, further suggesting that openness predicts creative achievement in the arts, but not in the sciences (Kaufman et al., 2016). Recent findings from research on the BTM conclude that only "plasticity" is positively and strongly associated with creativity (Feist, 2019). Similar results can be found in a study investigating Minecraft (a computer game) creativity and the openness in BFM through regression analyses (Shaw, 2022).

"Person" acts as a mediator between the brain and the manifestation of creative thinking and behaviour, and this mediator changes in accordance with the cognitive functions and behaviour exhibited in different scenarios and tasks (Feist, 2010; Kaufman & Sternberg, 2019). Different personal influences, such as personality or motivation, are reflected in students' actual performance, while their cognitive factors and performance, in turn, influence

their “person”. This thesis supports Csikszentmihalyi’s (1996) view that creativity is an integrative construct that can be observed and measured through student performances of cognitive functions on assigned learning tasks, regardless of individual influences that vary with the assessment context. While the influential factors represented by “person” are not independent of the whole construct of creativity, they are not considered as the best indicator of what students know and can do in relation to creativity and, as such, will not be pursued in this research.

1.2.2 Product

Product refers to the result or outcome of the creative process generated by the person. There are several ways of measuring the value of a creative product. One way is by assigning a numerical value to an individual’s eminence by counting the sum of an individual’s creative achievements (e.g., the number of awards and patents) (Colangelo et al., 1992; Simonton, 1980, 1998). Another is to have the individuals themselves self-assess their achievement using a Likert scale such as the Creative Behaviour Inventory (CBI) (Hocevar, 1979; Rodriguez et al., 2023) or the Creative Achievement Questionnaire (CAQ) (Carson et al., 2005), perhaps at multiple time points. Items that consider creative achievement in these assessments, such as awards, patents, publications, etc., are more likely to be used in assessing gifted or talented students and are not appropriate for defining and assessing creativity in the general student population.

An alternative way of measuring the creative product is to use domain experts, teachers or a peer review process and rate the product using a Likert scale that describes a

creativity continuum. The Consensus Assessment Technique (CAT) is the most typical rating scale used to measure a creative product or evaluate the degree of creativity of products or responses, by having appropriate domain experts or teachers using a 5-point scale based on their domain expertise and understanding of creativity (Amabile & Pillemer, 2012). The philosophy of the CAT allows it to be flexibly applied to different task scenarios and disciplinary domains. For instance, Kaufman and his colleagues (2021) used the CAT to measure students' creative production used in verbal and mathematics tasks, to investigate the relationship between creativity and traditional academic outcomes.

Despite the universal application of the CAT, core problems regarding its implementation cannot be ignored. A rater-dependent approach requires a clear understanding of the levels of raters' expertise as there are no specific criteria or a rubric for rating creative products (Cseh & Jeffries, 2019; Glăveanu, 2020; Glaveanu et al., 2020; Myszkowski & Storme, 2019). Varying levels of expertise or understanding of creativity can lower the inter-rater reliability, meaning that it is more likely that different teachers will rate the same performance of the same student differently (Reiter-Palmon, 2020). In short, improving the consistency of rating performance would require clearly articulated standards (describing what it is that characterises performance at increasing levels of creativity – the construct) provided to teachers in a way that enables those with variable experience and/or expertise to be trained to recognise and measure creativity in their daily teaching consistently against the standards.

1.2.3 Process

Process refers to the procedure used by the person to develop the product. The process of creativity involves a combination of conscious and unconscious (subconscious) steps, and its evaluation centres around the observable thinking skills of the process. The following tests have been used widely to measure the process of creativity across domains and grade levels.

Divergent Thinking Tests (DTT). The history of measuring the creative process dates back to Guilford's era in the1950s, when divergent thinking is dominant as a representation of the creative process (Guilford, 1950, 1959, 1966). According to Guilford's idea of creativity, divergent thinking can be assessed by three primary indices: fluency (the number of different answers), flexibility (the types of answers) and originality (the frequency of specific answers). There are four typical items that characterise divergent thinking tests include: unusual uses of common objects, instances of common concepts, consequences of hypothetical events and similarities between common concepts. Numerous divergent thinking tests have emerged that are based on Guilford's theory, such as the Kogan and Wallach Creativity Test (KWCT) (Wallach & Kogan, 1965) and the well-known Torrance's Test of Creative Thinking (TTCT) (Torrance, 1966b, 1969; Torrance & Ball, 1984).

The measurement of creativity has long focused on measuring divergent thinking, leading recent research to focus more on convergent thinking in a conscious attempt to increase the value of measuring both ways of thinking, and the broader view that the construct may align better with real-world creativity. Along these lines, test for Creative Thinking-Drawing Production (TCT-DP) measures creative synthesis by asking respondents to complete a drawing with a set of graphic elements (a dotted line, a half circle, a point etc),

where responses are evaluated against the unconventionality of the drawing, the number of graphic elements used and the connections between elements (Lubart et al., 2010). In the Evaluation of Potential Creativity (EPoC), divergent-exploratory thinking and convergent-integrative thinking are measured concurrently by asking students to generate many ideas from a stimulus, but then produce only one solution by integrating several elements (Barbot et al., 2011; Barbot et al., 2015; Lubart et al., 2013).

Remote Associates Test (RAT). Mednick (1962) advanced a different perspective on assessing the ability to think creatively and developed the Remote Associates Test (RAT). He suggested that individuals can generate unusual ideas because of the structure of the associations in their memories. The associative hierarchy can be measured by asking people to find the target words that are weakly related to the words provided in the item. More recently, the Associative Fluency Task (AFT) has started to gradually replace the RAT. By requiring respondents to generate lists of unrelated words, AFT overcomes the disadvantage of having only one correct answer to each RAT question (Acar & Runco, 2014; Beaty et al., 2014). As a proxy for creativity, semantic similarity is quantified by the probability of occurrence of a given word in each context. Applying computational methods to derive quantitative measures (e.g., Semantic Laten Analysis and Semantic Network) of semantic distance in AFT, associative thinking has been experimentally verified to be highly correlated with divergent thinking (Beaty & Johnson, 2021; Beketayev & Runco, 2016; Benedek et al., 2017; Hass, 2017; Kenett, 2018; Kenett et al., 2014; Kenett et al., 2023).

Although the use of the computation method makes it available to everyone on the internet, the tasks in AFT are too far removed from the classroom situation to be of general

value. It is unrealistic currently to assess creativity using tools and stimuli that are unrelated to the learning content. Also, the score of semantic similarity in AFT provides minimal feedback about what it is the individual students have produced as the manifestation of creativity, making it difficult for use in teaching and learning. More importantly, the nature of semantic analysis limits the observation of diverse forms of creative expression (e.g., visual, and auditory forms). This limitation makes it difficult for the AFT to be used in authentic classroom assessments.

1.2.4 Measurement issue

Traditional norm-referencing processes have been applied to the assessment of creativity (e.g., tests for creative process and products) and associated analyses of results and to give meaning to the scores by comparing them to the scores of a normative group of peers (Fishkin & Johnson, 1998; Plucker & Makel, 2010; Said Metwaly et al., 2017). In norm-referencing, scores are typically represented as percentile ranks showing the percentage of a reference group at or below the score. For example, a very creative score could be a student scoring at the 99th percentile or “at or above” 99 per cent of the norm group. That is well and good, but the results (score) are only meaningful in the context of the test, tasks or set of tasks. One problem with such a referencing model is that the results provide minimal feedback to teachers of what aspect of creativity the student has displayed and, as such, the results have little value in informing teaching and learning of creative thinking nor potentially improved student performance.

An effective measurement on creativity should provide direct feedback that can be

used by teachers to inform teaching and learning (Barbot et al., 2019; Skiba et al., 2010; Sternberg, 2015). In a scoping review of 51 research articles on creativity assessment in K–12 education, one dominant theme that emerged is the importance of defining standards for creativity assessment that are tailored to the context of the K–12 classroom, because they can provide informative feedback that teachers can use to design useful and effective pedagogies to promote the development of student creativity (Bolden et al., 2020). It is believed that applying standards-referenced assessment models to build a measurement rubric, with appropriate standards of performance (referred to performance standards), is imperative for creativity education (Barbot et al., 2019; Barbot & Reiter-Palmon, 2019; Henriksen & Mishra, 2015; Henriksen et al., 2020; Lucas, 2022; Mishra & Henriksen, 2013). In response to issue of measuring "product" using CAT, having articulated performance standard make it more likely that teachers with varying knowledge and experience will be able to make a consistent judgement on student performance. Such measurement rubric has the benefit of providing teachers with more meaningful and direct feedback about what students know and can do (to show growth in creativity) by referencing student performances to performance standards and locating them on the developmental continuum (the measurement rubric).

Beghetto (2019), McClure and Jaeger (2020) pointed out that the difficulty of using criterion-referencing assessments for measuring creativity is that creativity is determined by unexpected responses, unlike content-oriented classroom assessments, where the criteria for success are often more clearly defined. However, the arguement in this thesis is that building a measurement rubric for creativity is possible if essential and observable indicators of creativity can be identified and selected as components for the purpose, while characterising

the performance of students at different levels in each of the components as perfomrance-level descriptors. In this case, the outcomes from student peformance against tasks and items can be used to locate students on the measurement rubric by aligning the performance on the closest standard that best reflects the level of performance of the student on the task. Producing such a measurement rubric can be informed by drawing on existing measures of creaitivty, along with international frameworks of creative thinking (ACARA, 2016b; Brookhart, 2013a; OECD, 2019).

1.3 Research Aims and Research Questions

1.3.1 Research Aims

The purpose of this research is to build and validate a measurement rubric for creativity that, when used by teachers in a classroom, provides the basis for identifying the level of creativity of individual students in terms of how much of the construct they have demonstrated on a task (or a series of tasks), measuring the growth of students' creativity and providing teachers with evidence of the impact of their teaching and students' learning on creativity. The intention of assessing creativity is not to compare students' proficiency; rather, it is to detect and locate each student's current level of creativity along a clearly defined developmental trajectory.

The procedure for creating such a measurement rubric is based on the measurement theory approach outlined by Tognolini (2018). First is to operationally define the construct of creativity in a way that can fit the high school context. It must be stressed that the intention is not to arrive at a universal definition of creativity but rather one that fits the classroom

contexts that are the focus of this thesis (i.e., general high school, high school science and high school physics). The definition also needs to cross cultural boundaries, to be relevant in both Australia and China, and perhaps elsewhere. Second is to choose the components that represent the construct to give the explicit meaning of the definition. Third is to develop descriptions of performance for each incremental level of each component to show growth. Once the rubric has been constructed, a process of evidence collection is carried out to build a case to support the argument regarding the extent of the efficacy of the measurement rubric. This is key to its validation. The validation process begins with a review and evaluation of the measurement rubric by conducting a sequence of focus groups (designed to collect content and construct validity evidence). The next step is to design and evaluate appropriate domain-specific assessment tasks in physics as part of the evidential case to validate the practicality of the defined rubric for measuring creativity in high school physics (again, with content and construct validity evidence in mind). While this thesis is focused on assessment within the classroom environment, the process is consistent with large-scale, high-stakes assessment development (such as PISA), where definitions, language, assessment tasks, and processes are more widely socialised, standardised and validated for use in multiple jurisdictions (OECD, 2019, 2023a, 2023b; Schleicher, 2019).

1.3.2 Aims and Research Questions

Following are the aims of the research.

1. Define creativity and its components and develop a measurement rubric for measuring student creativity in high school classrooms.

2. Build an evidential case to support the efficacy of the measurement rubric as a measure of creativity in both generic and domain-specific contexts.
3. Demonstrate a process that Chinese Year 10 physics teachers can follow to develop tasks that enable the performance of students to be located on the measurement rubric.

To achieve these aims, the following research questions were designed to guide the research.

1. What is an appropriate measurement rubric to measure incremental growth along the defined components of creativity? (first research aim)
2. To what extent does a generic measurement rubric for creativity enable the measurement of growth in creativity across disciplines? (second research aim)
3. To what extent does a generic measurement rubric enable the measurement of creativity in the physics domain in the Chinese school setting? (second research aim)
4. To what extent can physics teachers in Year 10 classrooms in China develop assessment tasks that produce evidence that can be used to locate students within a generic measurement rubric for creativity? (third research aim)

1.4 Significance of the Research

The research is expected to make several significant contributions to the assessment and measurement of creativity. Modern measurement theory and standards-referenced assessment models are used to underpin the measurement rubric developed through this

research. In contrast to existing measures of creativity, this approach allows for provision of meaningful information on what it is the students know and can do in relation to creativity, beyond the value of the currently used norm-referenced measures. Measurement rubrics that describe in broad terms what growth in performance looks like can provide teachers and students with a better understanding of what it is that students know and can do with respect to the construct being measured (e.g., creativity). But importantly, measurement rubrics lay out a path for progression in the construct to help students and teachers on the continuing learning journey. By knowing where individual students are located along the measurement rubric and monitoring their growth along its developmental continuum, appropriate guidance can be developed to facilitate further learning. There are also advantages for students, who can get immediate feedback relating to where they are located on the construct and what they must do to get to the next level of performance. This research addresses the problems identified with current creativity tests and instruments in the earlier section of this report in that the rubric helps improve the understanding of creativity of those that use it and is therefore more useful to teachers, curriculum developers and policy developers.

In addition, significant benefits occur through teacher involvement in building, reviewing, validating, and using the measurement rubric. First, teachers can increase their understanding of creativity through discussion of the measurement rubric. This process provides teachers with an alternative view of assessment and measurement. Involving them in the process of designing appropriate assessment tasks to provide evidence relative to a creativity measurement rubric that can be reused is believed to be a new experience for Chinese physics teachers, as they generally tend to develop rubrics for each specific

assessment task. A generic measurement rubric that can be applied to different subject domains to measure creativity can further help teachers get a comprehensive picture of student creativity, as creativity demonstrated in different disciplines or subject domains can be located on this generic rubric. Involving teachers in building a measurement rubric for creativity can be generalised to measure other constructs that have previously been difficult (or impossible) to measure.

The process of developing and validating a measure in the way explicated in this thesis has the potential to be generalised – beyond the classroom, beyond just this research context and beyond the construct of creativity. In fact, it can be adapted for use with other 21st-century skills. The more effort that goes into validating the measurement rubric and standardising the tasks or situations, the greater the level of confidence that can be ascribed to the results from the process.

1.5 Structure of the Thesis

This thesis is divided into eight chapters. The current chapter outlines the rationale and context of the research. It articulates the important role of creativity in educational policies and national curricula around the world, and the urgent demand for extending the results from attempts to assess creativity into actionable outcomes in the classroom. Tests and assessments currently used to measure creativity are critically reviewed in terms of their reporting structures. The chapter includes a list of the aims of the research and a delineation of the associated research questions to be addressed to meet these aims.

Chapter 2 introduces the theoretical framework of modern measurement theory and

standards referencing that underpin the methodological approach used to govern the construction and validation of a measurement rubric for creativity. In general, the process of establishing and validating a measurement rubric follows the method developed by Tognolini (2018), which is appropriate for measuring constructs that are perceived to be difficult to measure, such as creativity and other 21st-century skills.

Chapter 3 illustrates the process of applying standards-referenced assessment methodology to assess creativity using a measurement rubric. Synthesising the research literature in relation to creativity develops the definition of creativity and its components that are essential and observable in classroom settings. The performance standards that, through performance-level descriptors, describe the incremental growth of performance relative to each of the components of creativity, are developed with referenced to the Revised Bloom's Taxonomy, Structure of Observed Learning Outcome (SOLO) Taxonomy and Associative Theory of Creativity to produce a draft measurement rubric of creativity. This draft measurement rubric is validated and improved in subsequent chapters.

Chapter 4 describes the rationale for using focus groups and teacher development workshops in the validation process. The thesis includes three validation phases involving three focus group (FG) discussions and two teacher development workshops. Qualitative data produced through these validation phases is analysed through qualitative content analysis. The analysis process is described in this chapter.

Chapters 5 to 7 present the results and validity evidence from the three focus group discussions and two teacher development workshops. Chapter 5 presents a reflection on the feedback received from Focus Group 1 (FG1), which comprised participants from a range of

disciplines. Chapter 6 considers the results from the discussion of Focus Group 2 (FG2), where the participants all have a science background, but not necessarily a physics discipline background. Chapter 7 summarises the outcomes from Focus Group 3 (FG3), which comprised participants who are teaching physics in China. It is this group that took part in teacher development workshops.

Chapter 8 concludes the thesis by discussing the results in relation to the research questions and considers the limitations of the thesis. Directions for further research on creativity assessment to improve teaching and learning in creativity are proposed.

Chapter 2 Theoretical Framework for Measurement Rubrics

2.1 Introduction

This chapter presents the theoretical approach used to construct and validate a measurement rubric. The measurement rubric developed in this thesis differs from the common understanding of a rubric (test-based or scoring rubric). Unlike the traditional rubrics that are directly related to the scoring of a specific test, a measurement rubric defines a construct to be observed and then shows growth through an area of learning by describing increasing levels of cognition – what it is the student knows, understands, and can do in relation to the components that comprise the construct.

2.2 Measurement Theory and Traditional Rubrics

The conventions of measurement theory (i.e., the classical and modern test theories) are designed for large-scale assessment to accurately estimate variables or qualities across any context (Brookhart, 2003; McMillan, 2013; McMillan, 2014; Scates, 1943; Shepard, 2006, 2019). In Classical Test Theory, generally a rubric is a test-based tool used by teachers and learners to accurately evaluate the quality (understanding, knowledge and skill in an area of learning) of student performance on a specific assessment test or task during the scoring process (Mertler, 2001; Panadero & Jonsson, 2020). Such a rubric is also referred to as an assessment rubric (Dawson, 2015). The associated assessment activities are generally criterion referenced in nature as students' specific performance on a course, test, item, task, or assignment are assessed by referencing the response to specific criteria in the rubrics (Nitko,

1980; Popham, 1971; Sadler, 1987). Unlike rating scales or checklists that only include a set of criteria for assessing students' work, rubrics also provide performance-level descriptors for the criteria (Brookhart, 2013b; Brookhart & Chen, 2015). Assessment criteria describe what aspect of quality is being observed, while performance-level descriptors indicate learning outcomes at each level of performance. Rubrics with criteria and performance-level descriptors, also known as descriptive rubrics, are effective tools for transparently communicating assessment expectations to teachers and students. This transparency enables teachers to focus their instruction on what is important and worthwhile while monitoring student progress during teaching and learning (Brookhart & Chen, 2015; Jonsson, 2014; Jönsson & Panadero, 2016; Lane, 2013). Rubrics are used extensively for assessment purposes across all levels of education globally (Brookhart, 2018; Brookhart & Chen, 2015; Jönsson & Panadero, 2016; Panadero & Jonsson, 2020).

Rubrics can vary depending on their design and how they are implemented (Dawson, 2015; Panadero & Jonsson, 2020). There are two main classifications of rubrics: generic or task-specific, and holistic or analytic (Jönsson & Svingby, 2007). Generic rubrics describe student performance across a family of similar tasks, while task-specific rubrics outline the specific requirement (e.g., facts, concepts, procedures) for students to include in their response to a given task (Brookhart, 2018; Panadero & Jonsson, 2020). Holistic rubrics provide an overall description of performance where all the components are combined and the on-balanced judgement (score) is obtained by aligning the performance on the task with the most closely matching description. Analytic rubrics provide descriptions of increasing levels of performance on each of the separate components that comprise the construct. They

provide a score for each of the components separately, identifying strengths and weaknesses of the individual parts of the student performance (Brown, 2017; Brown et al., 2018; Jönsson & Panadero, 2016; Mertler, 2001; Moskal, 2000; Nitko, 2001; Panadero & Jonsson, 2020).

2.3 Modern Measurement Theory and Measurement Rubrics

The measurement rubric developed and used for assessing creativity in this research is based on modern measurement theory derived from the Rasch Model (Andrich & Marais, 2019; Bond et al., 2020; Bond & Fox, 2007). The Rasch Model maps constructs onto linear continua, with equal units that can be adjusted from a set origin. In educational measurement, such linear continua have been called "developmental continua", "learning progressions", "progress maps", "construct maps" or "road maps" when each unit represents a level of performance with descriptors of what students know and can do in relation to the construct observed (Bennett, 2015, 2018; Black et al., 2011; Masters & Forster, 1996; Shepard et al., 2018; Wilson, 2018; Wilson & Sloane, 2000). Tognolini (2018) defined measurement as the "process of assigning a number to a performance to represent a location with respect to the developmental continuum underlying the performance that indicates how much of the construct being assessed is present" (p. 2). A rubric built using the tenets of modern measurement theory is referred to in this thesis as a measurement rubric. It is construct-oriented and focuses on describing a learning progression as an independent and unidimensional "ruler" to be used to locate student performance along the developmental continuum that represents the construct.

The process of giving meaning to student achievement at different performance levels

used in the measurement rubric is based on standards-referenced assessment theory. Although standard-referenced assessment relates closely with the philosophy of criterion-referenced assessment, it notably leverages teachers' qualitative judgements to work towards the specific standards to achieve satisfactory levels of comparability within an educational realm (Sadler, 1987, 2005). A standard is a definite level of a construct that describe what students know and are able to do. These descriptors of different levels of performance, which are cumulative to indicate growth of the construct to be measured, are referred to as performance standards (Bennett et al., 2012; Tognolini, 2018). A rubric built on a set of predetermined performance standards enables descriptors summarising a student's quality of performance or degree of achievement at a level to be made without reference to the performance of others; it can be applied to tasks of different types and does not require complex statistical analysis for interpreting the results (Sadler, 1987, 2005; Suskie, 2010). A standards-referenced assessment model aligns well with modern measurement theory and enables a measurement rubric to be operationalised in terms of a developmental continuum that describes a construct (in this case, creativity) through increasing levels of performance standards. A measurement rubric, therefore, allows evidence of student performance obtained from different assessment tasks, but related to the construct, to be measured and monitored over time by referencing the performance to the performance standards.

A generic measurement rubric is one that defines performance levels of a construct to be measured using general language that applies across domains or disciplines. Therefore, a generic measurement rubric is well suited to assessing a construct such as creativity, where the construct itself is sometimes embedded across domains or disciplines. In contrast, a

domain-specific measurement rubric defines a construct and growth using language specific to the discipline or domain. A domain-specific measurement rubric can be developed to align with the generic measurement rubric by contextualising the generic performance standards using the language and descriptors that are consistent with the domain, or discipline, but retaining the relativity of the performance levels described in the generic measurement rubric.

A measurement rubric can be analytic or holistic depending on the purpose of the assessment strategy. Analytic measurement rubrics are advantageous in providing formative feedback about student performance by presenting the standards separately for the different components, while holistic rubrics are helpful in providing an overall descriptor of performance at a specific level of performance (Brookhart, 2013b; Brookhart, 2018; Brookhart & McMillan, 2020; Dawson, 2015; Sadler, 2009).

The next section describes the theoretical approach used in this thesis to construct a measurement rubric.

2.4 Theoretical Approach Used to Construct Measurement Rubrics

A measurement rubric comprises a definition of the construct, its observable components and the associated performance standards (Bennett et al., 2012; Tognolini, 2018). The procedure for building such rubrics is consistent with those for building a developmental scale, progression framework or learning continuum that describes the typical order of learning development in relation to the knowledge, skill and understanding associated with a learning area (Black et al., 2011; Brookhart, 2013b; Masters & Forster, 1996; Scalise & Wilson, 2011; Wilson, 2009, 2018; Wilson & Sloane, 2000). Alignment between the

construct, components and performance standards supports the assessment result to be interpreted directly regarding the constructs observed, which links assessment with the curriculum and pedagogy (Lane & Tierney, 2008; Scalise & Wilson, 2011; Wilson, 2018; Wilson & Sloane, 2000).

Developing measurement rubrics for 21st-century skills, particularly creativity, is more complicated compared with rubrics for a specific area of learning with clearly defined goals. This is because the learning continuum for a generic construct spans multiple domains of learning (Care, Kim, et al., 2019). The definition and description of a generic construct may differ depending on the researcher, the underlying theory, and the purpose. However, it is important to have a clear understanding of the nominated construct so that each level of performance can be described accurately to comprise a learning progression. Here each step is demonstrated with the example of a measurement rubric used to measure graduate qualities at the University of Sydney (Tognolini, 2018).

Step 1: Define the construct to be measured. A construct refers to the quality observed in an area of learning, emphasising its meaning and understanding as the measurement goal (Lorge, 1951). In teaching and learning, a construct is generally understood to be a learning outcome or objective documented in educational curricula (Bennett et al., 2012). The University of Sydney defines nine graduate qualities (depth of disciplinary expertise; critical thinking and problem solving; communication; information and digital literacy; inventiveness; cultural competence; interdisciplinary effectiveness; an integrated professional, ethical and personal identity; and influence) by reflecting on the learning obtained through the generic curriculum offered across the university. One of the

graduate qualities – critical thinking and problem solving – is now used as an example to illustrate the process.

The definition of critical thinking and problem solving developed and socialised across the university is "questioning of ideas, evidence and assumptions in order to propose and evaluate hypotheses or alternative arguments before formulating a conclusion or a solution to an identified problem" (Tognolini, 2018). It is important to emphasise here that this definition and the subsequent components are only able to be generalised across the university because the process used to develop and validate the construct has used the university community as its entire population.

Step 2: Determine the components of the construct. Components emerge directly from the definition. They represent the construct and provide the target for the generation of evidence of student performance related to the construct. As an example, the definition of the construct of critical thinking and problem solving developed by the University of Sydney is the "questioning of ideas, evidence and assumptions in order to propose and evaluate hypotheses or alternative arguments before formulating a conclusion or a solution to an identified problem" (Tognolini, 2018). The components of this construct are: defining the problem or issue in context; critically questioning ideas; generating evidence and assumptions; creating and evaluating hypotheses or alternative arguments; formulating defensible conclusions; and creating best possible solutions to enable student performance relative to critical thinking and problem solving to be taught and measured within and across the university.

Step 3: Develop performance standards for each level of performance on each

component. Performance standards are further differentiated into performance-level descriptor that characterises the performance at each designated level. Each performance-level descriptor provides what it is students should know and be able to do to achieve each level of performance for each component by referring to relevant curricula and identification of characteristics of students at the nominated level along the developmental continuum. These levels are cumulative in the sense that, to achieve Level 3, students must achieve Level 2 plus "some more" of the component. The number of performance levels depends on the types of decisions that need to be made and the number of reliable distinctions that can be clearly described in terms of student performance. A measurement rubric generally contains four or five performance levels. Ideally, performance-level descriptors representing the performance standards should be explicit enough to enable unequivocal assessment of a component and hence the construct without necessitating specific items or instances of performance for reference, although such exemplars of performance should be made available where possible to enhance the accuracy of the decision making at each level.

Table 2.1 shows a generic measurement rubric for critical thinking and problem solving as an illustration of the outcome of the first three steps described above. It includes the definition, the components and their performance-level descriptors. Each component shows growth from low to high levels of performance, characterised by the performance described within each level. Each performance-level descriptor indicates what a student should know and be able to do at that level of the component. The procedures described above are applied in this research to build a measurement rubric for creativity. This measurement rubric is developed and presented in Chapter 3.

Table 2. 1 *Generic Measurement Rubric for Critical Thinking and Problem Solving Adapted from Tognolini (2018).*

Definition of critical thinking and problem solving:

Critical thinking and problem solving are the questioning of ideas, evidence and assumptions to propose and evaluate hypotheses or alternative arguments before formulating a conclusion or a solution to an identified problem.

Components:

- Definition of problem or issue in context
- Critical questioning of ideas, evidence and assumptions
- Creation and evaluation of hypotheses or alternative arguments
- Formulation of defensible conclusions and best possible solutions

Component	Level 0	Level 1	Level 2	Level 3
Definition of problem or issue in context	Does not offer a definition of the problem or issue, makes no attempt to situate the problem or issue in the context or to explain why it matters, makes no attempt to identify criteria against which to evaluate various solutions.	Provides a superficial definition of the problem or issue, shows that the problem or issue is situated in a context but is not yet able to show why the context matters or to identify criteria against which to evaluate various solutions.	Provides a useful definition of the problem or issue, shows that the problem or issue is situated in context, shows understanding in some detail of that context and can explain why it matters. Provides definitions of some key terms, can identify some	Insightful and articulate. Analyses and understands a context by consulting a suitably broad range of information sources, identifies and appropriately frames a problem or issue clearly in their own words, defines key terms, explains why this problem matters, sets out

			desirable features of various solutions.	criteria against which to measure various solutions.
Critical questioning of ideas, evidence and assumptions	Defers to received opinion without evaluating sources of information and without considering possible bias and error, does not recognise genuine expertise, does not identify or question methodologies used, is swayed by mere rhetoric, does not engage with evidence, does not consider other historical, intercultural or interdisciplinary perspectives.	Recognises that ideas, evidence and assumptions need to be examined, tries to identify possible bias and error, recognises that not every self-described expert has genuine expertise, identifies and avoids at least some fallacious rhetoric, provides evidence to justify conclusions.	Questions received ideas, evidence and assumptions, engages with the work of genuine experts, critiques fallacious rhetoric, engages in rational argument, assesses currently available evidence, shows awareness that new evidence may be discovered, shows awareness of differences in perspective.	Open-minded and intellectually rigorous. Critically examines received ideas, evaluates the credibility and methodology of authorities and experts, distinguishes sound reasoning from mere rhetoric, assesses currently available evidence, engages with competing views from various historical, intercultural and interdisciplinary perspectives, locates new evidence.
Creation and evaluation of hypotheses or alternative	Does not develop original hypotheses or arguments. Does not see how various hypotheses could be tested.	Recognises that current hypotheses and arguments may be suboptimal and ought to	Generates some new hypotheses and arguments, shows awareness of how they could be compared	Creative and judicious. Generates original hypotheses and arguments. Figures out how to test relevant

arguments		be evaluated, assesses at least some of the existing hypotheses and arguments.	and tested, carries out at least some of these tests.	hypotheses and arguments via reasoning, observation or experiment, carries out these tests, evaluates the results.
Formulation of defensible conclusions and best possible solutions	Does not offer any solution or conclusion, or dogmatically asserts a solution without being able to defend it. Does not identify and apply the relevant evaluative criteria.	Formulates an incomplete or unconvincing solution or conclusion and is not yet able to offer a convincing defence with reference to the relevant evaluative criteria.	Offers a solution or conclusion based on engagement with some of the relevant evidence, defends this solution or conclusion in light of relevant evaluative criteria.	Wise and decisive. Decides on the balance of the evidence, formulates conclusion or solution clearly in their own words, identifies the proper scope and significance of the conclusion commensurate with methods used, explains why this conclusion or solution is best when measured against relevant evaluative criteria.

2.5 Establishing Validity Evidence for a Measurement Rubric

A measurement rubric should be validated to ensure its effectiveness by collecting evidence of validity prior to the assessment activities (Brookhart, 2013b; Brookhart, 2018; Brookhart & Chen, 2015; Dawson, 2015; Jonsson, 2014; Kamei, 2019, 2020; Panadero & Jonsson, 2020; Wilson & Sloane, 2000). The process for collecting validity evidence used in this research is based on Kane's argument-based approach (Kane, 1994, 2006, 2016). The validity argument is synthesised using empirical evidence with the associated inference (theoretical evidence) to identify any strengths and weaknesses (threats). The validity argument may support the effectiveness of a nominated measurement rubric to be used in assessment practice or challenge it, which should lead to the revision of the measurement rubric to make it more valid and effective for assessment purposes.

Measurement rubrics designed to help improve student learning should involve classroom teachers in the validation process to increase the likelihood that the rubric is effectively used in classroom assessment. Classroom assessment is the process of collecting, evaluating and using reasoned evidence of student learning to diagnose students' strengths and weaknesses, monitor their progress, provide feedback and enhance learning (Brookhart, 2017a, 2017b; Brookhart & McMillan, 2020; McMillan, 2013; Towne & Shavelson, 2002; Wilson, 2018; Wilson & Sloane, 2000). Teachers should use their expertise and knowledge of the local context (specific learning environment, the taught curriculum and the relevant education policy), the nature of their students, content knowledge and effective pedagogies to collect evidence of student performance on the construct. It is the evidence that is used to locate the student on the measurement rubric according to the performance standard that best aligns with the student's demonstrated performance. Teacher involvement in the validation process can increase the likelihood that the measurement rubric is designed, aligned with the curriculum and used coherently in classroom teaching practices.

Designing assessment tasks is incorporated in the validation process to provide an evidential case for the efficacy of the measurement rubric. However, it is not usually conducted until after the measurement rubric has been constructed. As with measurement in the physical sciences, modern measurement theory requires a two-stage process where the measurement rubric is first developed and is then used to govern the construction of assessment tasks that are designed to collect evidence of student performance against the measurement rubric. Assessment tasks should provide evidence of a student's location along the developmental continuum (component) (Wilson, 2018). The tasks should align with the measurement rubric in a way that enables students of varying performance levels to showcase their knowledge and abilities (Tognolini & Stanley, 2011). The progressive nature of the components in a measurement rubric enables its use for both formative and summative assessment, which allows for a greater variety of assessment activities to be designed to produce substantial evidence that supports a reliable and valid judgement of the student's position against the performance standards (Black et al., 2011; Suskie, 2010; Tognolini & Stanley, 2011; Wilson, 2005, 2009, 2018).

2.6 Chapter Summary

This chapter clarifies the emergence and connotations of measurement rubrics by tracing back to modern measurement theory and standards-referenced assessment models. A measurement rubric describes what it is a student should know and can do at each level of performance. This is measured through the performance standards for the components that comprise the construct so that growth in relation to the construct can be observed. It enables teachers to collect evidence of student performance on the construct and locate them at different levels of performance by referencing the performance standards that best align with the performance.

The process of constructing a measurement rubric that has been illustrated through the sample of graduate quality – critical thinking and problem solving developed and socialised across the University of Sydney – is applied to the construct of creativity. In Chapter 3, a draft measurement rubric for measuring creativity is constructed followed the same process: defining the construct to be measured, determining the components of the construct and developing performance standards for each level of performance on each component.

Chapter 3 Constructing a Measurement Rubric for Assessing Creativity

3.1 Introduction

This chapter presents a process for building a measurement rubric to assess creativity. It starts with defining creativity and its components by reviewing available definitions of creativity predominant in the fields of psychology and education. Through a review of various creativity theories, appropriate components that are essential to the construct and can be observed in the classroom context are selected for inclusion in the measurement rubric for creativity. The final review addresses several learning taxonomies and Associative Theory of Creativity that help to describe and differentiate student learning and performance in the construct of creativity, thereby establishing performance standards to provide an image of what the growth of student creativity looks like in relation to the defined components of creativity.

3.2 Defining Creativity

While there is not a unanimously agreed definition of creativity in the literature (Childs et al., 2022; Mangion & Riebel, 2023; Plucker et al., 2004; Puryear & Lamb, 2020; Runco & Jaeger, 2012), one definition that has a degree of consensus is that creativity is the capacity to produce novel and appropriate ideas, solutions or products (Finke et al., 1992; Sternberg & Lubart, 1992a, 1992b, 1999). This definition includes examples of the types of creative achievement (ideas, solutions or products) and their characteristics (novel and appropriate). It is achievement-based and is referenced as being used in different educational assessment activities worldwide (Lucas et al., 2013; Vincent-Lancrin et al., 2019). However, such product-centric definitions imply a bias that individuals need to produce tangible products to demonstrate their creativity (Runco, 2014). This type of definition may not be

appropriate for describing students who are not yet capable of producing tangible creative achievement but whose creativity can be developed through schooling.

A second definition focuses on identifying crucial behaviours or actions in the creative process. For example, Wallas (1926) advanced four necessary steps indicative of the creative process: preparation, incubation, illumination and verification. This generic "Four-Stages Process" was later expanded and adapted to describe the creative process in different contexts, such as creative problem solving (Mumford & McIntosh, 2017; Mumford et al., 2012) and life-long social context (Csikszentmihalyi & Sawyer, 2015). Most teachers and educators who are interested in "creativity" in primary and middle school-aged students often associate creativity with the latter definition of the creative process rather than creative achievement in their teaching activities because of the level of competency of students (Barbot et al., 2015; Brookhart, 2010; Starko, 2013). However, creative achievement should not be ignored in schools when the assessment purpose is creativity, in which case it should be made explicit to accommodate students' level of creativity. If students are given the chance to create appropriate and novel work in school, their creativity can be recognised and evaluated. This consideration leads to the concept of "creative potential" that Barbot and his colleagues (2015) used to distinguish between and link the creative process and creative achievement.

Creative potential is a latent ability formed by a student's unique combination of various "ingredients", such as knowledge, cognition, conative characteristics and learning material, in the creative process with different ways of thinking (e.g., divergent thinking, convergent thinking) integrated to organise those various ingredients (Barbot et al., 2015; Lubart et al., 2013; Lubart, 2019). The term "potential" is used to indicate that having a comprehensive combination of the above influential ingredients can increase the likelihood of a student generating creative achievements if the student has a proper opportunity. However,

this likelihood is also affected by environmental factors and the match between the combination of ingredients and task requirements. This suggests that if the design of assessment tasks is adapted to the cognitive level and knowledge background of the students, and if the learning environment is positive and welcomes mind-challenging ideas or questions, students with creative potential are more likely to demonstrate their creativity by producing creative achievements.

The distinction, and simultaneously the connection between creative potential and creative achievement, is used to establish the definition of creativity in this thesis. This supports the intention that describes an individual's creativity in terms of both the process and subsequent achievement. Therefore, this thesis defines creativity as the capacity to generate creative achievements by exploiting an individual's creative potential.

3.3 Identifying the Components of Creativity

The next step is to identify the components of the measurement rubric. The components are independent of each other within the measurement construct and serve as rules or principles by which student responses, products or performances are judged. These components were not theory base and were selected as a consensus of various theoretical and researched suggestions for the definition of the measurement of creativity.

Problem finding. Problem finding has long been considered a significant indicator in identifying and assessing students' creative potential or creativity (Guilford, 1966; McClure & Jaeger, 2020; Mumford et al., 2013; Torrance, 1988). It aligns with one of the critical manifestations of creativity – sensitivity to problems (Guilford, 1950) – and fits the popular perception of student creativity, expressed as "always asking why" (Starko, 2013). Therefore, problem finding is selected as a component in the measurement construct of creativity and defined as discovering the purpose or subject to be communicated, and the discovery of

problems or needs in society (Getzels & Csikszentmihalyi, 1976; Starko, 2013).

Resource expansion. Along with the creative process, a series of strategies are used to collect and organise information against the defined problem (Gordon & Poze, 1980; Mumford et al., 2013; Prince, 1968; Ward et al., 1997). Resource expansion refers to selecting valuable and unexpected resources to satisfy one's purpose, which aligns well with the "association" of the creative resource (Lubart et al., 2013). In this component, the student uses varying strategies to find feasible resources or information. Resources are regarded as numerous kinds of information students use to generate an idea, such as their experience and knowledge.

Combination searching. Combination searching refers to synthesising selected resources by novel combinations, and shares a similar meaning to Guilford's selective combination (Guilford, 1959). Creativity expressed through this component requires students to search for as many methods as possible and then evaluate them to find the most novel and appropriate way to combine available resources. Several available instruments, such as the Test for Creative Thinking-Divergent Production (TCT-DP) (Urban, 2004; Urban & Jellen, 1996) and the Evaluation of Potential Creativity (EPoC) (Lubart et al., 2011; Lubart, 2019), indicate that resource expansion and combination searching can be observed through appropriately designed tasks. Resource expansion and combination searching contribute to other components that relate to the draft definition because they can illuminate the essential manifestation of creative potential.

Evaluation. Evaluation is one of the most important indicators of creativity, which requires students to be able to evaluate the extent to which the responses they have generated are worth the time and effort to pursue and extend (Guilford, 1950; Runco, 1991; Runco & Jaeger, 2012). Evaluation in creativity is demonstrated by students being able to select the most creative response from a range of answers they have produced and justify their choice.

Assessment of the evaluation component can be conducted in conjunction with the previous components in creative potential (i.e., problem finding, resource expansion and combination searching). The creative answers chosen by the student reflects a balanced judgement of their understanding of creativity, personal preferences and timeframe, and their consideration of the task context (Childs et al., 2022).

Creative achievement. Creative achievement shows the result of applying the creative process and is the outcome product of students' creative potential (Amabile, 1983b; Kaufman & Beghetto, 2009; Lubart et al., 2013). It has two well-known and accepted characteristics – originality and appropriateness (Amabile, 1983b; Guilford, 1950; Sternberg & Lubart, 1995). The general degree of students' creative achievement can be clarified by the Four-C Model (Kaufman & Beghetto, 2009, 2013) and the Triangular Theory of Creativity (Sternberg, 2018). The Four-C Model provides a developmental trajectory of individual creativity by distinguishing different kinds of creative levels, from Mini-c to Little-c, Pro-c and Big-c levels. The theory regards students' progress in learning as Mini-c and emphasises the creations being novel to the learners themselves. Little-c refers to creativity in everyday life and more focuses on creative outcomes that are valuable to others. Pro-c relates to creativity in professions, which is achieved after several years of professional and deliberate practice. Individuals at Pro-c level have more tangible achievements (e.g., award, patents, publications) that are assessed by specialised "gatekeepers" in a particular field, however, these achievements are not sufficiently remarkable to make significant revolution in the field or in the world to be located at Big-c level. Big-c reflects in world-changing creative productions, such as Picasso's Guernica Painting, Darwin's Theory of Evolution and Einstein's General Relativity. This model suggests that when measuring student creativity (i.e., Mini-c), the definition and standards should focus on what students have done to produce products that are novel to themselves. An alternative theory is the Triangular Theory

of Creativity, which distinguishes novelty into three levels: resistance to oneself, defiance of the crowd, and resistance to current thought. Resistance to oneself relates to the outcomes being new to the creators themselves; defiance of the crowd means the creative ideas or products are new to a cohort of peers; resistance to current thought is the highest level of novelty, which means the creative outcomes are considered new to contemporary society. As can been seen, resistance to oneself, as the lowest level of creativity in this theory, shares the meaning as Mini-c level that characterises student creativity. They suggest that the starting level of creativity should focus on the students themselves, encouraging them to come up with products that are novel to the students, avoiding peer comparisons and reducing discussion of societal value of their products.

Taking these components together, a draft definition of creativity adopted for this thesis is shown in Table 3.1, namely, it is the capacity to generate creative achievements by exploiting an individual's creative potential. Five components have been identified: problem finding, resource expansion, combination searching, evaluation and creative achievement.

Table 3. 1 *Draft Definition of Creativity and its Components.*

Construct	Component	Definition of the component
Creativity: a capacity to generate creative achievement by exploiting an individual's creative potential	Problem finding	Discovering the purpose or subject to be communicated, and discovering problems or needs in society.
	Resource expansion	Selecting valuable and unexpected resources to satisfy one's purpose.
	Combination searching	Synthesising selected resources through novel combinations.
	Evaluation	Identifying creative answers from multiple responses.
	Creative achievement	Producing novel and appropriate ideas, solutions and products.

The next section considers how the performance standard levels are conceptually arranged and constructed to describe the growth of each of the components and hence the growth in creativity.

3.4 Building a Performance Standard for Creativity

The Revised Bloom's Taxonomy, the Structure of Observed Learning Outcome (SOLO) Taxonomy, and Associative Theory of Creativity are used to build performance level standards to show the growth of each of the defined components of creativity. Dreyfus's (1980) Mental Model suggests five developmental stages for skill acquisition, ranging from novice to mastery. However, these developmental descriptions may not be appropriate to use for describing growth in creativity. According to Dreyfus's model, skill acquisition involves a transition or development from being able to use abstract (theoretical) information to using concrete experience through deliberate instruction and practice. This acquisition also reflects on the thinking process, from controlled awareness to automatic (intuitive) reactions. It means higher level learners can respond more quickly than lower level learners. Nevertheless, whether the information used is abstract or concrete (referring to the type of information) and how long it takes students to respond creatively (referring to the length of time to produce a response) could not differentiate between different levels of performance to show growth of creativity. According to Associative Theory of Creativity, growth in creativity is manifested in challenging students' mindsets and associating something seemingly unrelated to a stimulus to be creative. That is, creativity develops from recalling and applying materials that are closely related to a stimulus, to associating materials that are relatively distant from the stimulus so as to produce creative ideas and outcomes. This development process places little emphasis on the form of materials used (abstract or concrete) or the speed of students' response to the stimulus. The next section will further discuss the growth of creativity as

implied by Associative Theory of Creativity. Therefore, Dreyfus's Mental Model could not contribute to the performance standards of creativity as defined and developed within this thesis.

3.4.1 Theories used to Describe Growth in Creativity

Revised Bloom's Taxonomy. Bloom's Taxonomy of Educational Objectives (Bloom et al., 1956; Masia & Bloom, 1964) categorises learning objectives along a cognitive continuum from a teacher's perspective. The framework has six main categories: knowledge, comprehension, application, analysis, synthesis and evaluation. These levels indicate student progression, from simple to complex levels of cognitive functioning. Subsequently, a revision of Bloom's Taxonomy emphasises the dynamic conception of classification, using verbs to label the categories (Adams, 2015; Krathwohl, 2002). There are six categorised verbs: remember, understand, apply, analyse, evaluate and create. These "action words", which represent different levels of cognitive complexity from low-order thinking to higher-order thinking, have been widely used in developing lesson plans, suggesting learning and pedagogies, and assessment at all levels of education. In assessment activities, Bloom's verbs are frequently used to evaluate student performance by designing tasks that are consistent with the cognitive complexity of learning objectives. In this thesis, these words are applied to help elaborate on what cognitive complexity is required at each level of performance. These words also provide a guide for teachers to design appropriate assessment tasks for measuring creativity.

Structure of the Observed Learning Outcome (SOLO) Taxonomy. While the Revised Bloom's Taxonomy demonstrates the progression of cognitive complexity from the perspective of teaching and learning a nominated knowledge or skill, the SOLO Taxonomy considers the progression of outcome complexity by categorising the quality and depth of

students' actual outcomes in tasks through criteria of capacity, relating operation, and consistency and closure (Biggs & Collis, 1989, 2014). To be specific, the capacity refers to the information (cues and data) used in the outcomes, ranging from irrelevant or inappropriate (lower level) related and given in a stimulus, to related and hypothetical (higher level) to show incremental complexity of the outcomes. The relating operation refers to the relationship between a given stimulus and the learning outcomes to which students respond. The progression process of the relating operation is manifested in the following ways: outcomes at the lowest level are illogical; at the medium level it is concrete and question-specific; and at the highest level the outcomes are generalisable beyond the specific stimulus by relating to more abstract principles or concepts. The criterion of consistency and closure evaluates the consistency between information used and the relevant conclusions drawn by students. The more consistent the information used with the conclusions, the more likely it is that the outcomes will include more than one logical conclusion.

The progressions demonstrated by these three criteria, however, are unsuitable for describing growth in creativity because creativity encourages the use of different remote, and even seemingly unrelated information to stimulate novel insights and ideas, which is the opposite of what the capacity suggests. It suggests the consistency between the information used and the conclusions may not be a consideration in defining growth in creativity. The information used to generate outcomes can be given or hidden in the stimulus, or the abstract principles. It is the student's ability to flexibly associate and apply different resources (e.g., knowledge, skills, personal interest, stimuli) to a new task that is the key to creativity. As such, the progression shown in the relating operation may not align with what growth looks like in creativity.

However, SOLO Taxonomy does provide an evidence-based insight to define each performance-level descriptor by recalling and categorising students' performance in relation

to creativity in practice. This idea aligns with the theoretical approach to defining descriptors for each level of performance (see Section 2.3). It is especially helpful for constructs like creativity, which are yet to be well defined and whose developmental continuum remains relatively mysterious.

Associative Theory of Creativity. Associative Theory of Creativity is based on observing behaviours that lead to creative outcomes. According to Mednick (1962), creativity is the result of flexible associations between resources and stimuli. A flexible association requires students to have mastered the required elements organised in a complex network of associations against the stimulus. More flexible associations (or more flexible barriers) allow students to combine ideas and categories that are remote and unrelated together with stimuli. Students who have many experiences and resources (e.g., personal experience, knowledge, skills) with a given stimulus in a familiar setting, have a well-defined pattern of their responses, thereby are less likely to associate remotely with a given stimulus. This theory provides a theoretical basis for clarifying performance descriptors at different levels of the components to demonstrate growth in creativity. That is, differences in performance levels of creativity can be identified by the extent of complexity reflected in students' response association, specifically by determining the distance between the resources used and tasks (Hass, 2017; Kenett et al., 2014). Students who can use remote resources that reflect a complex network of associative thinking (or a flexible association) will be located at the highest level of creativity as defined by the measurement rubric.

3.4.2 Proposed Performance Standards for Each Component

Performance-level descriptors are used to describe the characteristics of growth in each of the components. They progress from low levels to high levels and are cumulative in nature. That is, the characteristics describing a Level 2 performance on a component would

include the requirement of a Level 1 plus more of the component. The performance-level descriptors are characterised by descriptive rather than judgemental (or evaluative) language and they describe the characteristics that are indicative of students with this "amount" of the component at that point on the developmental continuum.

3.4.2.1 Problem finding

The classical assessment of creativity scores students' responses through three creative characteristics – fluency (number of responses), flexibility (types of responses) and originality (frequency of responses) (Guilford, 1966; Runco & Jaeger, 2012; Torrance & Ball, 1984; Wallach & Kogan, 1965). These three characteristics reflect the cognitive development process of creativity, beginning with fluent thinking, followed by flexible thinking and ending with generating original ideas. The way that previous research weights fluency, flexibility and originality equally in creativity tests, to arrive at a total creativity score, leads to a practical problem in that they primarily rely on counts (or fluency) to indicate more of the construct of creativity (Hocevar, 1979; Kim, 2006; Plucker, 1999). Consider problem finding as an example. Producing many ideas simply or even making a long list in the same category does not indicate a high level of creativity (Perkins & Perkins, 2009). Inversely, students who can ask unfamiliar and meaningful questions could be recognised as creative even if they only produce one response. Several improvements have promoted the statistical operationalisation of these three creative characteristics to represent their significance and complexity, such as weighting them differently (from low to high) for summing, calculating only the sum produced by flexibility and originality, and using a ratio score calculated by dividing originality by fluency (Kim, 2006; Said Metwaly et al., 2017). Given the cognitive complexity represented by fluency, flexibility and originality, this thesis characterises originality as the performance for the highest level descriptor and locates students who can generate original ideas at the highest level of performance in problem finding.

Another practical issue is that some students cannot produce an original idea that can be used by teachers to identify creativity. Flexibility as an intermediate characteristic of creativity can be used to describe performance between fluency (low performance level) and originality (high performance level). Flexible thinking generally denotes the ability to look at a situation from many points of view or to generate many categories of responses (Lubart et al., 2013). Students who cannot generate a novel problem, but can put forward different kinds of problems, can be considered to have reached a higher level of creativity than those who can only discover problems in the same category.

In other words, if students are capable of generating novel and appropriate ideas or problems that exemplify the characteristic of originality, they would be viewed at the higher end of problem finding. Students who exhibit flexible thinking could be deemed to be located at a middle level of problem finding, and those whose performance is characterised by being able to make a list of responses in the same category (fluency) would be located at the lower performance level of the problem-finding component of creativity. To establish a performance standard for problem finding, the three creative characteristics are arranged at different levels of performance, from low to high, in accordance with their respective cognitive complexities. Simultaneously, Associative Theory of Creativity is employed to articulate performance descriptors at different levels, using appropriate Bloom's verbs to align with the associated cognitive complexity. This treatment of the three characteristics is applied to the components of resource expansion and combination searching. Table 3.2 shows the proposed performance-level descriptor standard for the component of problem finding.

Table 3. 2 *Proposed Performance Standard: Problem Finding.*

Level	*Performance-level descriptor*
4	Generate a previously unknown, unexpected problem.
3	Generate uncommon problems in similar categories or ask about common problems in remote categories.
2	Ask about common problems among similar categories or from similar perspectives.
1	Notice problems and ask about common problems within one category or from one perspective.

3.4.2.2 Resource expansion and combination searching

According to the above discussion on the cognitive complexity represented by the three creative characteristics, the performance descriptors from low to high levels in resource expansion and combination searching are characterised by fluency, flexibility and originality, respectively. In the component of resource expansion, students who can come up with novel and unfamiliar information related to the originality characteristics would be recognised as being at the highest level of creativity. Those who can collect resources from different but familiar categories, which reflects flexible thinking, would be located at a medium level of creativity. Those who can only collect resources within a single category, which implies the use of fluent thinking, would be at the lowest level of creativity. The same criteria are used to assign levels of performance to resource expansion as for problem finding.

A similar approach is used to differentiate between the levels of combination searching. That is, a high level of performance characterises those students who synthesise resources with a combination of previously existing strategies. A medium level of combination searching would be characteristic of those students follow a pre-existing approach in similar situations, to combining resources. The lower level would be demonstrated by those students who follow a pre-existing approach in an identical situation to combine resources. Tables 3.3 and 3.4 show the proposed performance-level descriptors

for the components of resource expansion and combination searching.

Table 3. 3 *Proposed Performance Standard: Resource Expansion.*

Level	*Performance-level descriptor*
4	Develop and assemble disparate resources that are novel and appropriate to the context.
3	Associate unfamiliar resources from similar categories or connect usual resources in remote categories.
2	Collect common resources across similar categories or disciplines.
1	List common resources within one discipline or category.

Table 3. 4 *Proposed Performance Standard: Combination Searching.*

Level	*Performance-level descriptor*
4	Create a novel and appropriate way that draws on previously existing strategies to synthesise resources.
3	Implement organised ways that use a combination of previously existing strategies to synthesise resources.
2	Follow ways previously existing within categories or domains in similar situations to combine resources.
1	Follow ways previously existing within a category or domain in an identical situation to combine resources.

The next section discusses the performance-level descriptors for the component of evaluation.

3.4.2.3 Evaluation

The evaluation component is demonstrated by students selecting the best answer and justifying their choice after they generate multiple answers to the open-ended questions. Levels of growth can be unpacked in the same way as the problem-finding, resource-expansion and combination-searching components. After generating multiple questions or resources in an appropriate task, the student can then be asked to choose the most creative

answer and give reasons to justify the novelty and appropriateness of the selected answer.

It is expected that students who operate at a high level of evaluation are able to identify and choose their most creative answer (Runco & Acar, 2012). Consequently, students who can choose the most novel and appropriate response with persuasive explanations would be located at the highest level of evaluation. If the explanation is not explicit enough to show the novel and appropriate aspects of the choice, it would be located at a middle level of creativity. The lowest level of evaluation is characterised by those students who choose one of the appropriate responses with a conventional reason. Table 3.5 shows the proposed performance standard for the evaluation component.

Table 3. 5 *Proposed Performance Standard: Evaluation.*

Level	*Performance-level descriptor*
4	Choose the most novel and appropriate response with a persuasive explanation.
3	Choose a relatively novel and appropriate response with a reasonable explanation.
2	Choose a relatively novel and appropriate response without clear reasons.
1	Choose an appropriate response to the context with conventional reasons.

The final component, creative achievement, is considered in the next section.

3.4.2.4 Creative achievement

Different domains have their preferable manifestations of creative achievement (Glăveanu, 2018). However, there are similarities across these creative achievements when distinguishing their extent of novelty (Weisberg, 2006). These commonalities are used to summarise and describe growth in the creative-achievement component of creativity. Various forms of creative achievement can be exhibited by producing tasks that require students to produce plans, solutions, paintings, handicrafts and creative writing.

This component is different from the previous component (evaluation), which focuses

on process elements. Creative achievement focuses on the outcome of the process. The general standard established in the measurement rubric provides teachers room for subjective evaluation based on their expertise. The level of students' creative achievement is dependent on the extent of novelty and appropriateness as measured by teachers using Amabile's Consensus Assessment Technique (CAT) (Amabile, 1983a; Amabile & Pillemer, 2012). This considers only the personal influence of creativity (Blamires & Peterson, 2014; Kaufman & Beghetto, 2009; Starko, 2013). The best evidence of whether the outcome is novel to the creator might come from a history of students' everyday activities captured on video and stored in a database, or through teachers' professional judgement based on the knowledge of their students and the characteristics of the way they approach problem solving (Starko, 2013).

If the achievement is a copy or simulation of an existing product or outcome, the student would be located at a lower level of creative achievement. If the achievement is a conventional combination of something existing and it is resolved and contextualised, the student would be located at a middle level of creative achievement. If the outcome shows novel aspects, and is appropriate and contextualised, the student would be demonstrating a performance standard at a high level of creative achievement. Table 3.6 shows the proposed performance standard for the component of creative achievement.

Table 3. 6 *Proposed Performance Standard: Creative Achievement.*

Level	*Performance-level descriptor*
4	Create an idea, solution or product that is novel, appropriate and contextualised.
3	Generate an idea, solution or product that shows novel aspects and is appropriate and contextualised.
2	Develop an idea, solution or product that is a conventional combination of something existing and/or is mostly resolved, practical and contextualised.

1	Make an idea, solution or product copied or restated from existing outcomes that is appropriate and contextualised.

The next section accumulates the performance-level descriptors and presents them in a framework that, because it is predicated on a developmental continuum (growth), is called a generic measurement rubric (see Section 2.2).

3.5 Proposed Draft Measurement Rubric for Creativity

3.5.1 A Generic Draft Measurement Rubric for Creativity

There has been significant discussion about whether creativity can only be demonstrated within a discipline (discipline-specific or domain-specific creativity) or whether it can be measured more generally (generically) as a construct (Baer & Kaufman, 2005; Barbot et al., 2016; Lubart, 2019; Weisberg, 2006). Domain-specific instruments embed the task and items into measurement contexts within the discipline or domain, while task and tests based on theories and tools that are not domain-specific (different questions can be used in different contexts) are referred to as generic. For instance, the Creativity-Science Assessment Test (C-SAT) (Ayas & Sak, 2014; Sak & Ayas, 2013) was created on the basis of a generic theoretical framework (Guilford, 1956; Torrance, 1966a; Torrance, 1988) and then applied within the contextualisd domain of scientific creativity (Klahr, 2003; Klahr & Dunbar, 1988). After developing the EPoC to measure the generic construct of creative potential (Barbot et al., 2011), several EPoC batteries were subsequently built to measure the same construct within specific domains (Lubart, 2019), such as the EPoC in Science (Vries & Lubart, 2017; Vries & Lubart, 2019).

As creativity is one of the generic “21st-century” skills, its measurement rubric should be designed so it can be used across different domains or disciplines. This draft measurement rubric is generic, meaning that it is not specific to any particular field and can be applied to different domain assessments. Therefore, the definition of creativity and its components,

along with their performance standards, should be generic to enable their interpretation and application in various practical assessments.

The domain generality of the creativity measurement rubric is supported by the Amusement Park Theoretical Model of Creativity (APT) and Weisberg's Cognitive-Analysis Procedure. The APT uses the metaphor of an amusement park to describe the relationship between generality and specificity in creativity. When walking around an amusement park, visitors begin by navigating the parks' distribution at the most general entrance, followed by the general thematic areas and more entertainment programs or facilities. Similarly, associating and using general creativity-related skills to a task is a starting point of creativity, gradually progressing to higher domain-specific levels (applying disciplinary-specific skills) to solve problems (Baer & Kaufman, 2005; Barbot et al., 2016; Kaufman & Baer, 2005). The APT is widely accepted as the theoretical foundation for describing and measuring student creativity in specific domains based on a generic framework for the construct of creativity (Huang et al., 2017; Sak & Ayas, 2013). It creates a bridge between generality and specificity when describing levels of performance in the components that comprise creativity. The starting point implied by APT is that the basic thought processes of creativity are generic (i.e., they subsume different contexts or domains). The argument is that, irrespective of the domain, the components and growth within components are the same (general). However, it could be that specific domains require additional skills to deal with problems in other situations. After an in-depth analysis of the creative process of generating outstanding products in varying fields, Weisberg (2006) highlighted the commonalities of the cognitive process underlying different creative outcomes and built the Cognitive-Analysis Procedure with the combination of top-down and bottom-up processes.

The APT model and Weisberg's procedure are used to guide the construction of a measurement rubric with these characteristics. Generic measurement rubrics describe growth

in each of the components, independent of the domain. These are referred to as generic-measurement rubrics. If it is necessary to describe the performance descriptors using terms specific to a domain or discipline, these are called domain-specific measurement rubrics. For example, a generic measurement rubric would have performance-level descriptors that do not refer to a discipline or subject. It describes growth in relation to the construct of creativity. If the descriptors reference a discipline or domain (e.g., science or mathematics) then they are referred to as domain-specific measurement rubrics, and if they directly reference a task within a discipline, they are referred to as task-specific measurement rubrics. The forms of expression of creativity, basic skills and necessary knowledge can differ across disciplines; however, they share the commonalities of the components of the generic measurement rubric. If it is necessary to build a domain-specific rubric to enable the convenience of measurement of creativity in classroom practice, a domain-specific measurement rubric can be derived from the generic measurement rubric in the disciplinary context.

3.5.2 An Analytic Draft Measurement Rubric for Creativity

The performance-level descriptors showing growth in different components of the definition of creativity are integrated as a draft measurement rubric shown in Table 3.7. The measurement rubric for creativity starts at Level 0, without any performance descriptors to characterise the situation, in that sufficient evidence has not been provided at this stage to place the student at Level 1 or above. This measurement rubric is analytic to provide teachers with more formative feedback for classroom-based assessments. Assessment tasks should be designed to enable the students to provide evidence of where they are located on the draft measurement rubric irrespective of the discipline or subject, because the focus is on the performance descriptor and the task within the various disciplines is written to enable the student to demonstrate the performance level within the discipline.

Table 3. 7 *Draft Generic Measurement Rubric for Creativity.*

Definition of creativity: Creativity is the capacity to generate creative achievement by exploiting an individual's creative potential.

Components:

- Problem finding: discovering the purpose or subject to be communicated, and discovering problems or needs in society.
- Resource expansion: selecting valuable and unexpected resources to satisfy one's purpose.
- Combination searching: synthesising selected resources through novel combinations.
- Evaluation: identifying creative answers from multiple responses.
- Creative achievement: novel and appropriate ideas, solutions and products.

Component	Level 0	Level 1	Level 2	Level 3	Level 4
Problem finding		Notice problems and ask about common problems within one category or from one perspective.	Ask about common problems among similar categories or from similar perspectives.	Generate uncommon problems in similar categories or ask about common problems in remote categories.	Generate a previously unknown, unexpected problem.
Resource expansion		List common resources within one discipline or category.	Collect common resources across similar categories or disciplines.	Associate unfamiliar resources from similar categories or connect usual resources in remote categories.	Develop and assemble disparate resources that are novel and appropriate to the context.

Combination searching	Follow ways previously existing within a category or domain in an identical situation to combine resources.	Follow ways previously existing within disciplines or domains in similar situations to combine resources.	Implement organised ways that use a combination of previously existing strategies to synthesise resources.	Create a novel and appropriate way that draws on previously existing strategies to synthesise resources.
Evaluation	Choose an appropriate response to the context with conventional reasons.	Choose a relatively novel and appropriate response without clear reasons.	Choose a relatively novel and appropriate response with a reasonable explanation.	Choose the most novel and appropriate response with a persuasive explanation.
Creative achievement	Make an idea, solution or product, copied or restated from existing outcomes, appropriate and contextualised.	Develop an idea, solution or product that is a conventional combination of something existing and/or is mostly resolved, practical and contextualised.	Generate an idea, solution or product that shows novel aspects and is appropriate and contextualised.	Create an idea, solution or product that is novel, appropriate and contextualised.

A lesson in science provides an example. To assess the component of problem finding in the lesson context of gravitation, teachers could show students an image of astronaut spacewalking as a stimulus and ask them to raise multiple questions that are novel to themselves and appropriate to the stimulus. Evaluation could be assessed at the same time by asking the students to choose the most creative answer as well as giving reasons for their choice. A further question, perhaps asking about gravity, could be constructed to assess the component of resource expansion. Students would respond by collecting multiple resources that fit the requirement in the question of being novel and appropriate in relation to gravity. Relevant resources would be made available to the students. The students are then asked to combine the resources in a variety of ways to solve the problem in a novel and appropriate way. For example, different machine parts could be listed as the stimuli in the question and students could be asked to combine some or all of them in various creative ways to create a dynameter.

The above hypothetical example describes the process by which a teacher could create stimuli to elicit evidence for each of the components. These stimuli should be designed to allow for different levels of performance to be elicited, according to the performance-level descriptors, so that the responses to the task provided by the students can be used to locate the students at the performance-level descriptor for each of the components that best aligns with the evidence provided by the students in their response to the task.

The last component is one that seeks an overall evaluation of a particular idea,

solution or product. Exercises like finding a novel solution to deal with water pollution can be provided as tasks, with students responding by developing a solution. Teachers locate the student on this component of the measurement rubric by using their professional judgement to align the response of the student to the performance-level descriptor that best aligns with the response. As these components are indicative of the construct of creativity, the various levels of performance, described in terms of performance descriptor levels, can be used to capture the "image" of what it is the student knows and can do in relation to creativity as captured in the generic measurement rubric.

Although some of the components are similar to the procedure of creative problem solving according to the Osborn-Parnes Model of Creative Problem Solving (CPS) (Isaksen & Treffinger, 1985; Parnes, 1981; Treffinger et al., 2003), they are not identical. The CPS consists of four specifying components (understanding the challenge, generating ideas, preparing for action and planning your approach), each of which is accompanied by phases as possible actions or tools to be used for solving the problem creatively. Problem solving is one of the essential reflections on student capacity. The student will be recognised as creative if the solution around a problem is creative (Treffinger & Isaksen, 2005). The capacity to solve problems creatively is often an important manifestation of creativity in the domain of sciences (Chumo, 2014; Hu & Adey, 2002; Sak & Ayas, 2013; Xin, 2019; Xu, 2018; Zhang, 2004). However, creativity is more than creative problem solving and embraces any form of creative expression (e.g., a unique painting that expresses a student's feelings and

ideas through a combination of different materials and drawing techniques). In addition, those who cannot come up with a creative solution might demonstrate that they have creative potential. Thus, the components are used to measure and locate creativity rather than just focusing on solving problems.

3.5.3 The Draft Measurement Rubric for Creativity Aligns with Educational Documents

The idea of incorporating commonalities of creativity and their varying levels of performance as incremental growth into a generic measurement rubric is not a novel attempt and can be found in the Organization for Economic Co-operation and Development (OECD) series of studies assessing creativity from a large-scale assessment perspective (Lucas et al., 2013; Vincent-Lancrin, 2015; Vincent-Lancrin et al., 2019). Through a thorough analysis of learning materials in schools (e.g., assessment tools, portfolio of pedagogical activities and exercises, and students' portfolios), OECD defines creativity as a generic skill that applies to a range of thinking and actions to produce something original and appropriate in a given context, with the process and product being the two components of the assessment rubric for creativity (Vincent-Lancrin et al., 2019). The above definition and rubric provide a solid foundation for the conceptual framework of creative thinking in the Programme for International Student Assessment (PISA) (OECD, 2019). The focus of PISA is on creative thinking as the cognitive aspect of creativity and includes three components: generating diverse ideas, generating creative ideas, and evaluating and improving

ideas. Students will be assessed through creative expression in written and visual forms, and problem solving in scientific and social contexts.

Creativity is a broad, multidimensional construct that encompasses a range of psychological factors (Guilford, 1950; Lubart et al., 2010; Mednick, 1962) that can be categorised into domain-relevant skills, creativity-relevant skills and task motivation, according to Amabile's Componential Model (Amabile, 1983b; Amabile, 2018; Amabile et al., 1996). Creative thinking, associated with creativity-relevant skills, has been long recognised as the primary representation for creativity (Barbot et al., 2011; Guilford, 1966; Torrance, 1988). Underneath the inextricable link between creative thinking and creativity, their assessment items (e.g., in the Torrance's Test of Creative Thinking and Guildford's Creativity Test) are similar, and student performance is differentiated through the cognitive depth represented by indicators of fluency, flexibility and originality. The creativity of school-age students is usually measured and represented by their performance on creative-thinking tests, the results of which are used to examine their creativity in relation to other influences, such as domain knowledge in science and mathematics (Huang et al., 2017), and the teaching and learning environment in which Science, Technology, Engineering, Arts and Mathematics (STEAM) programs are applied (Ozkan & Umdu Topsakal, 2021). Given the similarities between creative thinking and creativity assessment, the generic measurement rubric for creativity in this thesis was developed with reference to the research findings and curriculum requirements regarding the description of creative thinking and creativity, thereby increasing the likelihood of its application in

classroom assessment.

The Australian Curriculum, Assessment and Reporting Authority (ACARA) also focuses on its process and product of critical and creative thinking when building a generic learning continuum (ACARA, 2016b). This learning continuum organises four interrelated elements (identifying, exploring and organising information and ideas; generating ideas, possibilities and actions; analysing-synthesising-evaluating reasoning and procedures; and reflecting on thinking and processes) with incremental performance requirements for students at different grade levels. While this learning continuum has not been used as an assessment rubric to measure students' critical and creative thinking, it reflects specific learning goals for fostering and developing creativity in schooling. The measurement rubric for creativity developed in this research is consistent with the above generic learning continuum, as well as OECD's framework of creativity, in that a generic rubric or framework concentrating on the process and product can demonstrate student performance and learning progression in creativity during the process of teaching and learning.

The alignment between the measurement rubric for creativity developed in this thesis and the educational policy is also reflected in the selected components for the construct of creativity. The measurement rubric for creativity has similarities to the selected components of the other generic frameworks described above, as well as to the description of learning objectives for specific learning domains. For example, the Australian Curriculum: Science tags several specific learning areas in science as recommended ways to develop critical and creative thinking in students (ACARA,

2016a). These content descriptions include the skill sets of posing questions, making predictions and inferences, and solving problems, as well as inquiry skills that involve planning, selecting information and evaluating the information to generate pathways, solutions or conclusions. In the Standard of the Physics Curriculum for Senior High Schools in China, student creativity can be nurtured by allowing questioning when forming physical conception, by encouraging new solutions in the introduction of scientific thinking, and by providing opportunities for using new materials and technologies in physics experiments (Ministry of Education of People's Republic of China, 2017b). In addition, Brookhart (2013a) contextualised her understanding of a creative student into formative assessment and created a rubric for creativity in writing assignments to provide students with effective feedback on their learning. Brookhart believed that creative students could put things together in new ways, and included components of various ideas and contexts, various sources, combining ideas and communicating something new in the rubric.

Table 3.8 summarises the components of creativity selected in the three educational documents mentioned above, as well as in the measurement rubric for creativity developed in this thesis. The similarity to the components selected suggests that the proposed measurement rubric for creativity generally aligns with the educational objectives and could be applied to identify and measure student creativity in teaching practice.

Table 3. 8 *Selected Components of Creativity in Three Educational Documents*

Educational documents	Creativity Measurement Rubric	Australian Curriculum: Science	Chinese Standard: Physics Curriculum	Rubric for Creative Writing (Brookhart)
Components	Problem finding	Posing questions	Questioning and asking	Variety of ideas and contexts
	Resource expansion	Selecting appropriate information	Utilising different materials to design physics experiments	Variety of sources
	Combination searching		Combining materials with digital technologies to design new experimental approaches.	Combining ideas
	Evaluation	Evaluating evidence	N/A	N/A
	Creative achievement	Generating pathways, solutions or conclusions	Creative problem solving	Communicating something new

3.6 Chapter Summary

A draft generic measurement rubric is proposed for measuring creativity. It provides a common and clear description for teachers and other educators regarding what is intended to be measured (the definition of creativity), what aspects of performance are measured regarding creativity in assessment (the components) and how to locate this performance on a developmental continuum in creativity (the

performance standards). This measurement rubric for creativity has been developed based on a series of creativity theories, learning taxonomies and educational policy to improve its alignment with the schooling context and applicability in teaching practice. This is a draft of the measurement rubric with a strong theoretical foundation.

The next chapter explains the process involved in validating the draft generic measurement rubric for creativity. It is part of the construction process of the rubric expanded in this chapter.

Chapter 4 Validation Process: Draft Measurement Rubric for Creativity

4.1 Introduction

This chapter uses focus groups (FGs) and teacher development workshops to generate evidence in support of the validity of inferences drawn from the draft generic measurement rubric for creativity developed in Chapter 3. Four phases are designed to produce answers to the research questions, as shown in Table 4.1. The first focus group (FG1 – making judgements about the appropriateness of the measurement rubric across subjects) consists of participants from a variety of disciplinary backgrounds. The second focus group (FG2 – reviewing the draft generic measurement rubric through the science lens) involves participants with a background in science teaching or research. The third focus group (FG3 – validating the extent to which the generic measurement rubric could be used in Year 10 physics) includes six Chinese high school teachers in physics to participate in the discussion. The same Chinese physics teachers also participated in two teacher development workshops to develop tasks based on the measurement rubric as evidence in support of the validity of the measurement rubric. The first two phases, FG1 and FG2, in this research were conducted in New South Wales, Australia, and the final two (FG3 and two workshops) in Guangdong, China.

Table 4. 1 *Overview of the Evidence Collection Process for Validation.*

Research Question (RQ)	Phase (P)	Method
RQ2: To what extent does a generic measurement rubric for creativity enable the measurement of growth in creativity across disciplines?	**Validation Phase 1:** Validate the draft generic measurement rubric for creativity by involving educators across different disciplines	Focus Group (FG1)
	Validation Phase 2: Validate whether the draft generic measurement rubric for creativity can be used for the domain-specific context of science	Focus Group (FG2)
RQ3: To what extent does a generic measurement rubric enable the measurement of creativity in the physics domain in the Chinese school setting?	**Validation Phase 3:** Validate whether the draft generic measurement rubric for creativity can be used within the physics context	Focus Group (FG3)
RQ4: To what extent can physics teachers in Year 10 classrooms in China develop tasks to produce evidence that can be used to locate students within a generic measurement rubric for creativity?	**Validation Phase 4:** Design some assessment tasks to measure Year 10 student creativity in the physics domain	Teacher Development Workshops

4.2 Unified Validity and Validation in Assessment

4.2.1 Unified Validity

Validity in assessment is understood to be related to the interpretation of assessment results (Cronbach & Meehl, 1955; Messick, 1995) and requires the correct meaning of the assessment result and its entailing implications for action to be valid

(Cronbach, 1971). In the traditional validity framework, content, construct, criterion and consequential validity were the four main types of validity among hundreds of types of validity used to support or disprove inferences of intended argument. In essence, content, criterion and consequential validity contribute evidence supporting the trustworthiness of score interpretation, which echoes the underlying concept of construct validity (Messick, 1994). Considering the centrality of construct validity, the contemporary validity framework considered all validity as unitary to provide different sources of evidence from theory and data in supporting or refuting that a particular assessment measures what it intends to measure (Borsboom et al., 2009; Messick, 1989a, 1989b; Newton, 2012). Messick's framework was adopted in the Standards for Educational and Psychological Testing (American Educational Research Association, 2014), where validity refers to the extent to which the accumulated evidence from theory and data supports the score interpretation (inferences) for the proposed use of the test.

4.2.2 Validation as an Integrated Process of Evidence Collection

Validation is a scientific process of collecting multiple types of evidence to defend inferences made from a measure by integrating this evidence with logical inferences (Borsboom et al., 2004; Kane, 2006; Messick, 1998; Taylor, 2013). Under the overarching concept of unified validity, Kane (2006) proposed "validity arguments" to synthesise and compare results from empirical evidence with the expected (but also by anticipating the unexpected) interpretation and decisions, from

which their strengths and weaknesses are identified. An argument-based approach was subsequently formulated to provide professional guidance on the planning and evaluating validity arguments for assessment activities. It begins with a well-defined construct and associated performance criteria, followed by an explicit description of the intended assessment result and its interpretations (Kane, 2013). The American Educational Research Association, the American Psychological Association and the National Council on Measurement in Education have widely applied and advocated this approach to educational measurement and have developed standards for educational and psychological testing (2014).

The general procedure of building a measurement rubric used in this thesis theoretically shares the same understanding of unified validity and the argument-based approach, where it prioritises the role of the structural argument in the validation process to ensure the proposed measurement construct (i.e., the definition of the construct, its components and the associated performance standards) measures what it claims to measure – hence focusing on the traditional bodies of evidence commonly referred to as content and construct validity. The intertwined link between the validation and development process enables the use of validity evidence to improve and refine the measurement construct to provide more integrated validity of the measurement (Forster & Masters, 2010; Shepard, 2016). In response to the research questions, four validation phases were designed to generate and solidify the validity arguments to support the claims underpinning the draft measurement rubric for creativity. The purpose of the first three phases is to collect evidence, through

three different groups of educators, whether the draft measurement construct is well defined and easily understood, and explicitly describes the growth trajectory of the components of creativity as it is defined in this thesis. The measurement rubric is the primary underlying source for the subsequent design of assessment tasks used to produce evidence that can locate students on the measurement rubric. To further validate whether the measurement construct of creativity can be embedded in teaching practice and used to measure student creativity in a classroom setting, two teacher development workshops were held in which physics teachers were asked to develop tasks for use in their classrooms. The tasks were to be aligned with the creativity construct, and the resulting student responses to the tasks used as evidence to locate students along the measurement rubric and, in so doing, describe what it is they know and can do in relation to the components of creativity.

While most current validation studies pay more attention to test development and the validity of the proposed instruments (Chasteen & Scherr, 2020; Li et al., 2023; Rohr-Mentele & Forster-Heinzer, 2021), some studies focus on the interpretation of the measurement construct or rubric, as in this thesis. For example, McKenna and his colleagues (2022) conducted a validation study involving focus groups, expert panel reviews, interviews and testing results to support the validity of an early elementary writing rubric for kindergarten and first-grade students. Different validation studies embody the same nature of collecting different theoretical and empirical evidence to form validity arguments as evidence for or against the proposed use.

The next section describes the characteristics of the focus groups that have been used in four phases of the validation process.

4.3 Methods of Focus Groups and Teacher Development Workshops

4.3.1 Focus Groups

Focus groups (FGs) originated as a research tool in marketing research, beginning in the 1950s, and were generally used to gather consumer preferences and promote products. However, they have also been used extensively in social science (Merton et al., 1956) and educational research (Creswell, 2014; Merriam & Tisdell, 2015; Morgan & Hoffman, 2018). Typically, a FG involves an interactive and in-depth discussion on a nominated topic with a purposely selected group of people who share similar backgrounds or experiences, with an optimal group size of six to ten participants (Mishra, 2016). By sharing and reflecting on a focused topic, FG participants can refine their perspectives based on what they have heard (Hennink, 2014) and ultimately reach a shared understanding of the proposed topic. These descriptive inputs can provide researchers with characteristics of social and cultural norms and people's perceptions of an issue, from which potential and valuable topics can be explored for further research (Morgan, 1996).

In educational measurement and assessment research, detailed input from FG participants often involves logical arguments about the feedback, attitudes, understandings and reviews on a given topic (Brookhart & Nitko, 2019; McMillan, 2014; Miller et al., 2013). Since collecting structural arguments is the focus of

validating the measurement rubric for creativity in this thesis, tailored discussion questions were predetermined to guide the FG discussions to collect feedback and comments from FG participants regarding the chosen definition of creativity, its components and the associated performance standards (performance-level descriptors) developed and presented in the draft generic measurement rubric for creativity. The sharing of ideas and views among participants enable the identification of important factors to refine and assess the quality of the measurement rubric through an analysis of its various parts. In this way, the FG approach used in this thesis differ from traditional focus groups in which participants are usually invited to share their respective views on a topic freely. This modification in the FG methodology resembles the review groups that characterise the evaluation of items in a test development process (Haladyna & Rodriguez, 2013).

The next section describes the second data collection stage involved in the validation of the measurement rubric, teaching professional development workshops.

4.3.2 Teacher Professional Development Workshops

Teacher professional development is a continuing process that is intended to enhance educators' up-to-date and effective knowledge and skills to maximise the likelihood of quality education (Steyn & Niekerk, 2005), which can be achieved through several learning activities such as workshops, conferences, consultation, coaching, seminars and courses (Bernadine, 2019). The success of professional development activities generally requires the active participation of teachers in a

collaborative community bonding with a clear and shared objective that is highly relevant to their workplace (Dede et al., 2009; DuFour & Eaker, 2009; Fullan, 2007), which is supported by a multi-case study in the context of high school social studies (Powell & Bodur, 2019). Dalby (2019) created a research cycle for professional development from class observations, meetings and teacher interviews, in which relevant elements (knowledge of practice, knowledge for practice and knowledge in practice) could provide opportunities for knowledge sharing and reflection on practice.

The teacher development workshop is one effective method used when professional development is to "turn ideas into practices" (Joyce & Showers, 2002; Sariscsany, 1996). This is consistent with the requirement inherent in answering the final research question related to teachers designing assessment tasks to extend and apply the knowledge and skills to assess students' creativity. These assessment tasks, developed collaboratively in the workshops, will contribute to the evidential argument used to validate the measurement rubric for creativity.

The next section describes the methodology used in the various validation phases.

4.4 Samples and Procedures Used in the Validation Phases

4.4.1 Validation Phase 1: Validate the Draft Generic Measurement Rubric for Creativity by Involving Educators across Different Disciplines

Given the second research question – "to what extent does a generic

measurement rubric for creativity enable the measurement of growth in creativity across disciplines?" – the first focus group, with diverse disciplinary backgrounds, reviewed and commented on the draft generic measurement rubric for creativity. To meet the group characteristic of domain and discipline diversity, volunteers with experience researching or teaching in a high school context were eligible to express their interest in participating in FG1. When selecting FG1 participants, preference was given to those from different disciplinary backgrounds, given the intention to review the generic performance-level descriptors. FG1 comprised eight PhD candidates from the Sydney School of Education and Social Work at the University of Sydney, whose subject backgrounds included mathematics, languages (i.e., English and Korean), mathematics, engineering, arts, human society and its environment, physical education, science and technology.

Discussion questions for FG1 are presented in Appendix D. The questions focused on seeking feedback and reflection on the definition, the components and the draft performance-level descriptors (performance standards) constitute the draft generic measurement rubric for creativity. The answers to the questions provided the substance for the validity arguments used to support the extent to which the draft generic measurement rubric could be used across different disciplines to measure creativity. Significant feedback from FG1 participants was synthesised using qualitative content analysis and adopted to increase the quality and validity of the draft generic measurement rubric, which is presented in Chapter 5. The researcher acted as a moderator and interpreter to facilitate FG1's cross-interactions. Necessary

materials (including the rationale for the draft generic measurement rubric for creativity, information related to performance standards and standards referencing) were prepared in advance and shared in the meeting as the advance organiser for the discussion.

4.4.2 Validation Phase 2: Validate whether the Draft Generic Measurement Rubric for Creativity Could Be Used for the Domain-Specific Context of Science

Focus group 2 (FG2) was then conducted as the second validation phase to ascertain the extent to which the generic measurement rubric could be applied in the domain-specific context of science. While FG2 was conducted in much the same manner as FG1, the discussion questions (Appendix E) were slightly altered in a way that required participants to review the draft generic measurement rubric through a science-discipline lens. During the review of the generic measurement rubric for creativity, participants further commented on the extent to which there was a need for a domain-specific measurement rubric to measure student creativity in science classrooms. If the consensus was that, with minor or no modification, the generic measurement rubric could be used to measure creativity using science-oriented creative tasks, then the generic measurement rubric would be retained and used in the subsequent phases. If the results of FG2 suggested there was a need to rewrite the generic measurement rubric to make it more usable for measuring creativity in the science discipline, then a domain-specific version of the measurement rubric would be developed, which is internally consistent with the growth structure characteristic of

the generic measurement rubric. This would mean that, in practice, science teachers could use the domain-specific measurement rubric to measure creativity in their science classrooms, but that the performance-level descriptors would retain a direct alignment with those of the generic measurement rubric, even though the performance descriptors themselves would be written with a science orientation. This would allow a school to report holistically on creativity despite the nuanced differences in domain-specific rubrics employed by each discipline.

The criteria used to select the participants in FG2 were: (a) have at least one year's experience of teaching or researching in high school science classrooms and (b) have a sound knowledge of standards-referenced assessment. Two science-discipline participants from FG1 volunteered to also participate in FG2. This allowed the understandings related to the considerations and deliberations shared in FG1 to be accurately communicated to FG2 through the researcher and these two participants. There were six participants in total in FG2, including the three science teachers from FG1, one head teacher in science and two professors in science education.

4.4.3 Validation Phase 3: Validate whether the Draft Generic Measurement Rubric for Creativity Can Be Used within the Physics Context

The draft generic measurement rubric for creativity has been designed to provide a practical and effective school-based measure that can be used in different disciplines to assist teachers in recognising the development of student creativity in China. This thesis demonstrates the process of building and validating the

measurement rubric by using the discipline of science. In China, the science domains are divided into physics, chemistry and biology, which commence as different science domains in Year 7 (Ministry of Education of People's Republic of China, 2017a). Given the specific disciplinary setting of science domains in Chinese high schools, Phase 3 was designed to collect evidence, through a separate focus group (FG3), about whether the generic measurement rubric and the domain-specific rubric (if developed at the end of Phase 2) were appropriate for use within the domain or subject of physics in a Year 10 Chinese classroom. As such, FG3 added significant information to the validation argument.

The criteria of selection for FG3 were: (a) have at least one year of experience in teaching high school physics and (b) was teaching physics in a Chinese high school in 2022. Schools that focus on educational research and the development of students' capabilities were given priority and sent invitations for teacher participation in this study. Six physics teachers from provincial model high schools in Guangdong volunteered to participate in FG3. All volunteers had a master's degree in physics teaching. Materials such as the generic measurement rubric, introductions to modern assessment and creativity theory prepared for earlier phases were translated into Chinese and shared with the participants in FG3. The discussion questions (see Appendix F) were similar to those used in FG2. In addition, participants were asked to comment on whether a domain-specific rubric in physics was necessary for measuring student creativity in a physics class or if teachers could measure just as effectively using a generic measurement rubric. If the participants agreed that a

domain-specific rubric in physics was required to make evaluating student creativity in their class more valid, then a physics-specific rubric would be developed for use in Phase 4.

The qualitative content analysis of the FG3 participants' comments resulted in the finalisation of the draft generic measurement rubric, moving it from being a draft rubric to the final generic measurement rubric for creativity. As the participants of FG3 consistently favoured the use of a generic measurement rubric over a domain-specific rubric in physics, the final phase used this generic measurement rubric as the basis for designing and constructing assessment tasks appropriate for use in physics classrooms in China.

4.4.4 Validation Phase 4: Designing Assessment Tasks to Measure Year 10 Student Creativity in the Physics Domain

This phase considers the crucial role of assessment task quality for effective measurement. Alignment between tasks and performance levels allows student responses to be aligned with performance standards, making learning visible through the generic measurement rubric. While the assessment tasks are not part of the generic measurement rubric, they are designed to enable students, through their response to the tasks, to provide evidence about their performance and at the same time provide evidence that can be used to ascertain the extent to which the generic measurement rubric is fit for purpose – linking the assessment task and the collected evidence from student performance for the measured construct (Cizek, 2020; MacCann & Stanley,

2010; Margolis et al., 2016; Masters, 2013).

The second task for FG3 participants was to design some assessment items or tasks that, when answered by the students in their class, would provide evidence regarding their location on the generic measurement rubric. The participating teachers were further asked to provide any possible responses from students and align them to the measurement rubric since administering tasks was beyond the scope of this thesis (i.e., the research did not collect students' actual responses to the designed tasks). Even though FG3 participants were invited to join the upcoming workshops, half withdrew from this stage due to job-related factors. The study invited three replacement participants who met the selective criteria of FG3. These three new participating teachers were from provincial model high schools in Guangdong and had a master's degree in physics teaching. Therefore, six physics teachers were involved as participants to design and finalise assessment tasks that, when given to students, would provide evidence from their responses that could be used to locate the students on the measurement rubric.

Two teacher development workshops were designed to provide tailored professional support when designing tasks for evidence applicable to the generic measurement rubric for creativity. The first workshop served an introductory purpose to address potential challenges that may have related to their knowledge of modern measurement theories and understanding student creativity according to feedback from the previous FG3 discussion. Specifically, the participants were introduced to the relevant theories, including standards-referenced assessment, the composition of a

measurement rubric in general, its function and application for creativity. Test items in mathematics and science from the National Assessment Program – Literacy and Numeracy (NAPLAN) and the National Assessment program – Science Literacy (NAP – SL) in Australia were used to exemplify the alignment between the assessment tasks and their associated measurement rubrics. The workshop then presented several task design principles specifically for assessing higher order thinking (referring to analytical, evaluative and creative thinking according to the Revised Bloom's Taxonomy) (Brookhart, 2010) and asked participants to share their understanding of these principles. This workshop concluded with a discussion of existing assessment tasks of creativity in the literature and those proposed by early FG members with a focus on aligning items and the draft generic measurement rubric and the extent to which items met the task design principles. The introductory workshop was designed to provide the necessary knowledge base and design insights for developing appropriate tasks to measure creativity in the second workshop.

The second workshop was designed to focus on task development. Participants used their measurement and assessment knowledge to design and refine tasks that could be used to measure creativity in their classroom by focusing on the components in the generic measurement rubric. After determining the tasks, they were asked to provide possible student responses that demonstrated varying levels of performance creativity and link them to performance standards. The evidence for validating whether the designed rubric supports measuring creativity in physics lies in the assessment tasks and the linkage between responses and performance standards. The

workshop concluded with a discussion on the importance of creating a task-specific rubric, to support consistent interpretation of students' responses by teachers (Wilson & Scalise, 2006). If the consensus was that a task-specific rubric would make it easier for teachers to reference student responses to the performance descriptors in the generic measurement rubric, participants would provide example responses from hypothetical students at each level of performance against the tasks by referencing the generic measurement rubric. Student responses at different levels would then be characterised as performance-level descriptors to constitute task-specific rubrics that are aligned with the generic measurement rubric.

Two iterations of peer review of teachers' initial tasks allowed teachers to share ideas with each other and deliberate on the appropriateness of their tasks. Participants have the potential to develop teacher assessment literacy through these iterations of writing a task against a measurement rubric and defending their own ideas. Through these iterative processes, teachers' challenges in assessing creativity were better understood, leading to suggestions for tailored professional support in measuring creativity in classrooms.

4.5 Data Collection

Each of the focus groups and teacher development workshops in the different validation phases lasted two hours and were conducted online via Zoom due to Covid restrictions on face-to-face activities. The four phases of data collection were performed sequentially, as the analysis results in the former phase were directly

applied in the latter phase. All online activities, except for the first introductory workshop, were recorded for precise transcript and qualitative analysis of participants' feedback and perceptions of the draft creativity measurement rubric.

Taking the data collected in Validation Phase 1 as an example, the participants' responses to the discussion questions and their interactive statements during the group discussion were transcribed and analysed through qualitative content analysis as validity arguments to support the draft generic measurement rubric for creativity refined. The type of data collected and the approach to data collection and analysis in Phases 2 and 3 were the same as in Phase 1, except that the background of the participants was different. FG3 and the second workshop were conducted in the Chinese language. The data were translated into English by the researcher for analysis purpose. The data in Phases 2 and 3 were transcribed and summarised through qualitative content analysis to determine if the generic measurement rubric for creativity could be used in a science and physics context and whether a domain-specific rubric in the physics context was necessary.

The data collected in the second teacher development workshop included participants' personal descriptions of their designed tasks, peer feedback, defending statements and subsequent editing, and their perception of building a task-specific rubric for measuring creativity. These data provided evidence to construct an evidential argument to address the final research question regarding the extent to which Year 10 physics teachers in China could develop tasks that, when given to students, consistently produce evidence to locate students within a generic

measurement rubric for creativity. The first two types of data were obtained by transcription. The assessment tasks and the task-specific rubrics (if applicable) were submitted via email. Peer feedback, defence statements and improvements to several edited tasks reflected teachers' professional development in educational assessment and measuring creativity. Table 4.2 shows the type of data and the method data were collected in the second workshop.

Table 4. 2 *Data Sources and Collection in the Second Workshop.*

Duration	Content	Data collected	Data collection
10 mins	Researcher overviewed the task design principles and measurement rubric for creativity.	N/A	N/A
20 mins	Participants developed their initial tasks that enable the measurement of creativity against the rubric.	Initial tasks	Submitted via email
20 mins	First peer review: Participants shared their tasks with each other and provided comments according to the judgement criteria.	Task descriptions, peer feedback, defending statements	Transcription from audio recording
Break (10 mins)			
15 mins	Participants refined their tasks according to the peer-review feedback.	Second draft tasks	Submitted via email
20 mins	Second peer review: Same as the first peer review.	Task descriptions, peer feedback, defending statements	Audio transcription
10 mins	Participants finalised their tasks according to peer-review feedback.	N/A	N/A
15 mins	Group discussion: Was there a need for a task-specific rubric?	Comment and feedback	Audio transcription

4.6 Method of Data Analysis

4.6.1 Qualitative Content Analysis

As one of the main methods for analysing focus group data (Onwuegbuzie et al., 2009; Vaismoradi et al., 2013), content analysis (CA) refers to the process of using coding and categorising to examine data and explore words used in relation to their trends, patterns, frequencies, relationships, communicative discourses and structures (Grbich, 2007; Mayring, 2004, 2015). While typical CA, in a chiefly quantitative manner, focuses on quantifying qualitative data by counting the frequency of the occurrence of various categories in a dataset (Neuendorf, 2018), its variation, known as qualitative content analysis (QCA), emphasises the subjective interpretation of qualitative data (Hsieh & Shannon, 2005).

The purpose of several focus groups is to generate direct evidence of the validity of the generic measurement rubric to measure creativity and to improve the draft generic measurement rubric by critically reviewing participants' comments and feedback based on knowledge and evidence drawn from the literature. The purpose of two teacher development workshops is to collect evidence regarding the extent to which teachers can produce tasks that can be used to locate students on the generic measurement rubric (additional validity evidence). It is the participants' responses to the draft generic measurement rubric for creativity that provides the answers to the research questions in this thesis. The QCA is used to analyse the transcribed data from this perspective.

4.6.2 Procedure for Conducting QCA

Transcribed data in this research were analysed following the general procedure in QCA, including preparing, organising and reporting (Elo & Kyngäs, 2008).

4.6.2.1 Preparing

QCA starts by identifying relevant data for the purpose of analysis, which can accelerate the initiation of other steps in the data-analysis process. This process involves listening to the audio recording, reading relevant material and transcribing verbal data to build an intimate immersion and comprehensive understanding of the data.

In this research, the data were transcribed from focus groups and teacher development workshops to build an overall understanding of what was being provided by the respondents in support of the arguments for the validity of the data and ultimately the validation of the generic measurement rubric. Therefore, the analysis focused on the "manifest content", which refers to the words used or visible meaning in the data.

Transcribed data from focus groups and teacher development workshops were initially grouped according to their discussion questions (see Appendices D–F) and scheduled processes (see Table 4.4 as the preliminary data groupings for the second teacher development workshop). These groupings provided several overarching topics for further organising participants' comments. The overarching topics for focus groups comprises the definition of creativity and its components, and the performance

standards. The topics for teacher development workshops include assessment tasks and peer-review feedback. The researcher translated the data from the FG3 and teacher development workshops from Mandarin to English and shared these related analyses (Mandarin and English) with the participants to ensure the quality and accuracy of the translations.

4.6.2.2 Organising

Organising data involves deconstructing the data by coding valuable words or phrases and clustering relevant codes to create emerging categories and subcategories. Descriptive materials are broken down into small content units as codes. Subsequently, several cohesive codes are examined and regrouped as categories or themes to represent some paramount aspects of a complex phenomenon that participants talk about frequently at an in-depth level (Castleberry & Nolen, 2018; Sparkes, 2005).

In this research, participants' comments from focus groups were compared in each aspect to synthesise several discussion topics. Feedback was summarised to produce arguments to support the credibility of the draft generic measurement rubric for creativity (i.e., provide validity evidence). Feedback that did not support the validity argument of the current generic and domain-specific (if applicable) rubric were carefully considered by reflecting on the purpose of this research and referencing the relevant literatures so that necessary amendments were made to improve the validity of the measurement rubric.

Taking FG1 transcribed data as an example, different phrases and words

representing participants' understanding and attitudes during the discussion were italicised as codes. The following phrases were marked as cohesive codes as they revolved around the use of "achievement" in the draft definition of creativity.

I'm confused about the word "achievement" because it is not very specific.

Achievement is often used in the syllabus and reminds teachers of some mechanistic and conservative marking experiences.

Can we clarify what a creative achievement is?

Creative achievement is infinite.

These phrases were grouped under the "achievement" topic. This organising process produced a database of evidence that did not support the use of "achievement" in the draft definition of creativity. In response to this concern, a necessary amendment was made to "achievement" by elaborating different forms of achievements and referring its connotation to the descriptive phrase used in the literature.

4.6.2.3 Reporting

Reporting elements comprise the analysis process, the interpretation of the integrated topics and the analysis results (Sutton & Austin, 2015). Section 4.6.2 describes the analysis process, while Chapters 5–7 demonstrate several substantive and compelling topics and their analysis results for each validation phase. Each topic will be discussed in depth with direct quotes from empirical data to make the process of accessing conclusions transparent.

4.7 Ethical Considerations

Before commencing data collection, this research received the Human Research Ethics Committee (HREC) approval from the University of Sydney with Protocol Number 2021/774 (Appendix A). The data collection, analysis process and reporting of the study results strictly followed the research proposal submitted to HREC to guarantee the anonymity and confidentiality of all study participants. In FG1, the researcher sent an invitation letter to administrative officers at the Sydney School of Education and Social Work as intermediaries to contact potential participants. In FG2, the invitation was sent to academics and research students in SSESW and the Faculty of Science at the University of Sydney, the University of Technology Sydney and the University of New South Wales. A letter of invitation was sent to administrative officers in the Catholic Schools NSW (CSNSW) and the Australian Independent Schools NSW (AISNSW) to invite science teacher participation from schools. The potential participants in FG3 and teacher development workshops were contacted by passive snowball recruitment. Invitation information for this study was advertised in physics teaching interest groups on WeChat. These invitations described the purpose of the study, outlined what was involved in the participation if they agreed to participate and requested them to contact the researcher if they were interested in participating in the study. The above methods avoided the possibility of real or perceived coercion in recruitment, as participants' contact details were only available when they initiated contact with the researcher.

All participants received a Participant Information Statement (PIS, Appendix

B) and Participant Consent Form (PCF, Appendix C) and needed to sign and return the PCF to the researcher prior to the start of the study. The PIS states clearly that participation in this study is completely voluntary and that participants are free to withdraw from any sessions by informing the researcher verbally or via email.

To protect participants' confidentiality and their data privacy, participants were represented by the code to remove identifying information in this thesis. The code starts with the letter P as a participant and two digits for the participant number, followed by a focus group number (e.g., FG2). For example, participant 2 in FG2 is identified as P02_FG2. In this case, any subsequent use of participants' comments in this study was completely anonymous and no personally identifiable data were used in any of the transcriptions.

4.8 Chapter Summary

Validity evidence for a draft generic measurement rubric that describes growth in each component of the defined creativity construct was collected and coded. This supported a validation process through four phases involving three focus group discussions and two teacher development workshops. QCA was applied to analyse and synthesise the key topics and the associated statements from the participants involved in the validation process. These topics and statements provide evidence via validity arguments to support the validity of the draft generic measurement rubric for creativity. The results of this data collection for the generic focus group (FG1) in Validation Phase 1 are reported in the next chapter.

Chapter 5 Validation Analysis: Generic Focus Group

5.1 Introduction

This chapter presents the results from Focus Group 1 (FG1) as part of the evidential case to answer the research question of the extent to which the generic measurement rubric for creativity enables the measurement of growth in creativity across disciplines by involving educators across different disciplines. The result consists of discussions and reflections on three overarching topics based on the predetermined structure of the discussion questions (see Appendix D): the draft definition of creativity, the defined components and their performance-level descriptors that constitute the performance standards. The participants' comments were categorised into discussion topics based on the relevance of the content of the discussion, while their direct statements were quoted to support validity arguments for the generic measurement rubric for creativity.

5.2 Reviewing the Draft Definition of Creativity

The draft definition of creativity is "the capacity to generate creative achievement by exploiting an individual's creative potential". In this validation process, the participants in FG1 started by considering whether they agreed with the definition. They further discussed what they thought a more appropriate definition would be and how the draft definition could be improved. The overall structure of the draft definition was first discussed, followed by a forensic analysis of the meaning of

the words used in the draft definition.

5.2.1 The Overall Structure of the Draft Definition

The draft definition of creativity integrated both creative potential and creative achievement to convey the idea that students should be taught and allowed to think broadly (understood as creative potential) to generate creative achievement when the environment allows. This integration received positive feedback from FG1 participants. For example, Participant P05_FG1 said:

> I love the idea of [creative] potential and achievement as an art teacher. It reflects the journey of creativity and how we come into that as a capacity, and how we can build that in kids. Creative products can be very difficult, so I love how you've recognised the difference in those things.

Another similar response came from Participant P03_FG1:

> The way you've [the researcher] put the [creative] potential and achievement reflects the context in a teaching-and-learning setting. It really works for me…the distinction [between creative potential and creative achievement], and it is a big positive.

After receiving supportive comments about the general structure of the definition of creativity from participants in FG1, the discussion shifted to a thorough review of the definition's meaning to ensure that it provides a clear understanding of creativity from a teacher's perspective and that the wording is appropriate.

5.2.2 Meaning of the Words Used in the Draft Definition

Six keywords evolved during this part of the focus group session: capacity, individual, potential, achievement, creative and exploiting. The first five relate to the clarity of the definition. The term "exploiting" refers to the emphasis on active and constructive learning during the measurement process.

(1) "Capacity"

In the draft definition, "capacity" was originally used in reference to Weisberg's (2006) analysis of creativity. However, participants argued whether it was an appropriate word to describe student development in creativity. They believed that there was an essential difference between "capacity" and "capability" and suggested the use of "capability" because they were of the view that it better emphasised the learnability of creativity in the definition. Participant P03_FG1 believed using the term "capacity" would be problematic when considering a diversity of students:

> When thinking about every single student, you [a student] can have that creative potential, but you actually have to have the capacity to take that creative potential into something, which potentially rules out lots of people …

Supplementary explanations were added by Participant P04_FG1 to clarify the distinctive nature of these two terms:

> … like motivation and perseverance, they are important. Certain perseverance is inherent in creativity because it makes you continue to work … students' motivation can be low, but their capacity might be

> high. They are quite different things.

Participant P05_FG1 also commented on this:

> … the argument is that everyone has the capability, and that's about harnessing that and growing so that everybody has the capability somehow. That's what research shows us … the capacity is the innate opportunity to build that particular thing in an individual. When you're changing the word, you're changing what you're assessing.

The unified recommendation of the participants to use "capability" as an amendment to the definition indicates an alignment in the belief of learnability in creativity between teachers and the literature. Creativity has been widely considered a teachable generic construct that can be developed through school education (Beghetto, 2018; Kaufman & Beghetto, 2009; Lucas et al., 2013; McClure & Jaeger, 2020; OECD, 2019). Terms like "capacity", "skill", "ability", and "capability" are used commonly and interchangeably in different definitions. For example, Sternberg and Lubart (1999) used "ability" to define creativity; the OECD (2019) applied "skill" to understand creativity; "capacity" was used to describe creativity in the National Advisory Committee on Creative and Cultural Education (NACCCE, 1999), while creativity in the Australian curriculum is defined as a "capability" (ACARA, 2016a).

The learnable nature of creativity implies that students are not inherently limited to a fixed level of creativity. Creativity can be nurtured or hindered through environmental influences that a person encounters later in life. In this sense, "capability" was considered by FG1 participants as a more fitting term for the

measurement of student creativity.

(2) "Individual"

Several influential factors are at play in the systematic structure of creativity. The term "individual" was adopted to emphasise the selection of teachable and robust influences for personal development through schooling as a measurable component. It was not intended to restrict or control specific types of measurement in practice. Different types of assessment tasks are appropriate if they provide the opportunity for students to demonstrate their capability, and this can be referenced to the measurement rubric. However, using "individual" in the definition may mislead teachers into limiting themselves to individual assessment only, as commented by Participant P05_FG1:

> I would also take out the word "individual" because the other way to be creative is through collaboration. It is about harnessing the potential of others to employ you and build on your own capability. That's what the researchers are telling us, and I think it is also beautiful in the research. It's not just about the employment of my own creative potential, I need to employ the creative potential of others as well.

Collaborative learning, such as project-based learning, has been applied in art and STEM courses to enhance and measure student creativity (Anitha et al., 2018; Hanif et al., 2019; Storer, 2018; Wu & Wu, 2020). The feedback from taking out the word "individual" in the definition was fully supported by all participants as they consensually believed the freedom of the assessment approach could maximise the

efficacy of the measurement process using the measurement rubric. The fewer unnecessary restrictions on assessment in the definition, the broader the range of measures that can be considered and used in instruction. When applied to students, their responses can provide teachers with evidence that can contribute to a greater level of confidence in the measure that is produced. In response to the suggestions of the participants in the focus group, "individual" was removed from the definition to account for the diversity of learning and the integrity of the eventual measure.

(3) "Potential"

The participants endorsed the significance of creative potential in student creativity. However, the direct use of the term "potential" in the definition was problematic in that it could lead to confusion because it is non-representational and vague in meaning. It potentially obscures the meaning of creativity and potentially prohibits teachers from building an image of what students can do in relation to their creative potential. As Participant P08_FG1 questioned:

> [In order] to define the aspect of creativity or being creative in education, we should be very careful in identifying creative action here, like "creative potential". If using this concept here, I would ask how you define "creative potential", in what ways do we know that students can show their creative potential …

School teachers will be the ultimate implementors of measuring creativity in their classes. To do so, they will first have to understand the definition of creativity and its descriptors in the measurement rubric, then design appropriate assessment

tasks accordingly. They will then have to reference students' responses to the tasks to the performance descriptors defined in the measurement rubric. As emphasised by participant P04_FG1, definitions of any ability or measured construct should be concise and easy to understand for teachers, as long as the new definition is still meaningful and the change has not modified the meaning of the construct. Participant P01_FG1 advised replacing the abstract concept of creative potential with its specific components, outlined in the measurement rubric. This would eliminate any unnecessary terminological hierarchy and reduce the reading load for teachers. FG1 participants responded positively to this proposed amendment, finding it to be clear and easily understandable for school teachers, with the change not altering the construct. As a result, creative potential in the definition was amended to its measurable components while retaining the same overall defining structure of creativity. These components include problem finding, resource expansion, combination searching and evaluation.

(4) "Achievement"

"Achievement" is a general word used to describe any learning outcomes that students may achieve in different subjects and stages of learning, such as paintings, essays, songs, and solutions to a problem. It is used interchangeably with outputs and outcomes when describing the result of students' creative potential (Barbot et al., 2015; Lubart et al., 2013). Nevertheless, a definition with unrestricted requirements for the outcome could make teachers bewildered about what kind of outcomes should be focused on. Confusion was put forward by two participants, with one of their

comments (P06_FG1) shown as an example:

> Normally, we talk about creativity as the ability or capability to generate novel and appropriate ideas, not that achievement … when you say achievements, I'm confused by that word because achievement is not very specific. It could be absolutely anything …

Having incongruous feelings about the meaning of "achievement" is understandable for those teachers who have just started to think of teaching and assessing creativity in their classes. School teachers tend to think and instruct in their preferred and familiar ways, where learning outcomes are mostly knowledge-based and well described in their associated curricula and syllabi. However, as Participant P01_FG1 noted, the requirement to be specific is contrary to the myriad of possibilities that are characteristic of creative outcomes. Teachers should be mind-challenged and swift in their traditional ways of thinking if cultivating complicated capabilities like those described as 21st-century skills.

However, Participant P01_FG1 agreed to change the word "achievement" because it is often used in the syllabus and reminds teachers of some mechanistic and conservative marking experiences. Participants analysed this issue by discussing and clarifying diverse forms of creative achievement that could happen in their teaching practice, including but not limited to verbal ideas (e.g., verbal discussion and spoken poems), texts (e.g., essays, written plans, computer output and handwritings) and products (e.g., experimental products, plans, handicrafts, painting). There is no end to the list of creative expression. Still, throughout this process, participants felt much

clearer about what creative achievement looks like and suggested the provision of these examples in the definition and rubric for better understanding. Participant P06_FG1 supported this suggestion in the following way:

> I think discussing and listing various forms of creative achievement is quite clear to understand achievement here. It's better to put writing, handicraft, and painting under the product category, as "product" basically means a physical product. It can be separated from an idea, a plan, or a solution. I think there are lots of different kinds of outcomes that students can give us, and these are some of the examples …

As the discussion progressed, however, participants found that examples grouped under different forms of achievement overlapped in their connotations. For instance, "a plan and solution can be done in an essay form and the discussion as well", stated by Participant F05_FG1. "It really depends on what teachers are going to assess in terms of creativity and how to represent their achievement as an end goal."

As per the feedback received, the term "outcome" was used to replace "achievement" in the definition to remove the focus on grading and comparison during the measurement process. This change was made because both terms have a similar meaning referring to the result of the production process. To make more explicit what is meant by creative outcome, the phrase "ideas, solutions and products", which is commonly found in creativity literature, was retained to define creative outcomes. The participants were of the view that providing representative forms of creative achievement enables teachers to get a better understanding of

creative outcomes. At the same time, it gives scope for teachers to design any task that is appropriate for their classes so it assesses creative achievement.

(5) "Creative"

Participant P05_FG1 made the following comment:

> The definition is trying to explore what creativity is, and it's a little bit tricky and problematic to use the word "creative" or "creativity" in the definition when that's the thing that we're looking to define. It can make the construct of creativity harder to understand …

The issue arises from using the synonym "creative" in "creative potential" and "creative achievement" to define creativity. This concern was discussed by the participants in FG1, and several suggestions were made to address the issue. The first proposal was to completely remove the term "creative" from the definition of creativity. This was overwhelmingly rejected by the participants because this term highlights the significant characteristic of creativity from other capabilities. The second suggestion was to utilise special font styles such as italics, inverted commas and bolding to symbolise its own meaning in the definition. The significance placed on the term raised the query as to, "What precisely is meant by 'creative'?" The answer to this question led to the final and workable recommendation: replace "creative" with the language used in the literature. The participants believed this correction could eliminate any repetition in semantics and provide a precise definition of creativity. As novelty and appropriateness have been linked quite closely in the literature to creativity (Guilford, 1965; Sternberg, 1982), "creative" was replaced by

the terms “novel and appropriate” in the definition.

(6) “Exploiting”

FG1 participants were mindful of using words appropriately to encourage active learning rather than provide negative connotations of learning when discussing the definition of creativity. They felt uncomfortable when they encountered the term “exploiting” and found it conveyed negative connotations when applied to teaching and learning because one of its meanings is “to use someone or something unfairly for your own advantage” (Cambridge University Press, n.d.).

The use of “exploiting” was designed to harness two essential notions of creative potential and achievement in the draft definition, indicating creative generation from a series of cognitive processes to an eventually tangible outcome. The original intention was to convey a similar meaning as other synonyms in this thesis, such as “using” and “employing”. After explaining the original meaning conveyed by “exploiting”, participants recommended choosing other synonyms that imply positive learning. As Participant P01_FG1 declared:

> It’s [“exploiting” is] a negative connotation … if you give this definition to teachers or students, then you know it will raise something negative. You [the researcher] are not going to be around to explain the meaning of the words when teachers use the rubric.

And from Participant P05_FG1:

> It’s an exploitation of your individual potential to deliver the outcome, and that’s how we do it. But I think the consensus is that the word

> "exploitation" doesn't give that lovely positive, joyful phrase that we want to give to creativity's definition instead of sounding like oppression.

The participants in the focus group searched for appropriate alternative expressions to solve this problem. Any words with negative connotations or those that are not commonly used in teaching practice were excluded. For example, the word "utilising" has a similar problem to "exploiting", implying the use of something beyond its intended purpose; "leveraging" was believed to be inappropriate because it is normally used to describe tools or tangible materials. The discussion ended with an agreement on the use of "employing" because it implies active participation and sustained action. In this case, "employing" was selected to take the place of the term "exploiting" in defining creativity.

5.2.3 Conclusion regarding the Draft Definition of Creativity

In FG1, the extent of the consistency between the structure of the definition and teaching practice, and the appropriateness of the concepts and words used in the definition were discussed and examined based on participants' knowledge of creativity and understanding of their students. The adopted amendments aimed to clearly explicate the meaning of creativity by using direct, simple and positive words. According to the above analysis, the redrafted definition of creativity is "the capability to generate novel and appropriate outcomes by employing problem finding, resource expansion, combination searching and evaluation".

5.3 Reviewing the Draft Components of Creativity

5.3.1 Alignment between the Draft Components and the Definition of Creativity

To review the alignment of the draft components with the definition, participants in FG1 were asked to consider whether the draft components needed to be reworded or eliminated, and whether additional components needed to be included to better align with the description.

Participant P02_FG1 commented:

> … [the] alignment is very clear. All the components match the definition very well. They [these components] show us the full process of producing a creative outcome. I think they are all important and cannot be eliminated …

And from Participant P03_FG1:

> I agree. All components match the definition very well, as they are used directly in the definition.

The above comments received several "agrees" from other participants. Overall, FG1 found a good connection between the current components and the definition of creativity. After this positive feedback from FG1, these components were retained without any modification.

5.3.2 Appropriateness of the Draft Definition of the Components of Creativity

The participants of FG1 generally agreed with the definition of the components drafted by the researcher because "these definitions unpack the

components and creativity well", commented Participant P05_FG1. However, some participants in FG1 argued that the phrase "the discovery of the purpose or subject to be communicated" is unclear in the definition of problem finding. This generated a heated discussion around understanding "the purpose to be communicated". Another issue emerged with respect to the definitions of the components. Participants indicated that they found it difficult to distinguish between these components of creativity and those associated with normal ways of thinking. These two issues are taken up in the following session.

(1) "The purpose to be communicated"

In contrast to the conventional understanding of problem finding, the new phrase of "the discovery of the purpose to be communicated" in the definition was drawn from Starko's (2013) consideration of highlighting the effort to communicate so that school students would be allowed to use different forms of expression to convey their own voice. Nevertheless, the participants were perplexed when confronted with this newly added phrase, as Participant P05_FG1 stated:

> I am unsure about using the word "communicated". Do you mean communication is part of the defining of the purpose? Then it's about "communication", not about "being communicated" … Communication is inherent in exploring ideas, enquiry, and curiosity [that kind of] is harnessed in that finding stage. But this word ["communicated"] gives me the sense that they must communicate the idea, and they might not know what that is yet, like in that problem solving and that inquiry and

> that exploration. There is still coming together to [kind of] put that together and can still be communicated …

Participant P05_FG1's follow-up suggestion of using "subject to be explored" to include a broader range of observable performance for measuring creativity received both support and opposition. Participant P01_FG1 agreed with this terminology in preference to "communicated" or "communication" because:

> "Communicated" is a harsh word for me to use around grades and things like that, and that's not necessarily what you want them to do here. Communication to me is telling students that they must find a solution as opposed to what is a problem from my perspective.

Participants P02_FG1 and P03_FG1 disagreed with this suggestion when considering its implementation in their classes. Participant P02_FG1 said:

> If I consider the situation of students' collaborative work as a group when implementing any assessment for creativity, I prefer to keep "communicated" or "communication" in this definition … exploring is more like finding new things, but "communication" is based on their experience and knowledge. When students are communicating, they can explore further [more] levels of different category things. If you're saying "explored" for class, how can students in Year 7 at Stages 4 and 5 have a picture in their mind of what they are going to do about this problem finding? Maybe teachers will encourage them to communicate with each other.

And participant P03_FG1 said:

> It's more purposeful when it's "communicated" rather than "explored", so there is potentially something there to do with the observable nature of what you're looking at because it's hard to observe exploration. I'm not sure of what the resolution would be …

The impasse prompted the participants to rethink whether the first half of the definition – either about "be communicated", "communication" or "explored" – must be included. When trying to delete the controversial part of the definition, FG1 participants found that "the discovery of social problems or needs" can satisfy the purpose of communication as well as exploration more generally. Participant P03_FG1 summarised this as follows:

> I think the definition can be a single statement. The addition of the "and" verb in the definition means that you've got two parts, which then means I can have the discovery of purpose but not the discovery of problems and still fulfil the problem … if using the word "or", you just not really settled on the problem itself …

Other wording issues around the first half of the definition also produced some further changes to the definition. The word "purpose" used in the definition was criticised by Participant P04_FG1 because "it really puts a value on it [a problem]; it is something quite definitive". The word "subject" in this contentious phrase conveys an ambiguous meaning that could be interpreted as school subjects or discussion topics, stressed by Participants P06_FG1 and P07_FG1. They suggested combining

two halves of the definition by replacing “subject” with “topic” or “idea” and deleting the rest of this phrase.

To concatenate all the advice and make a single statement of problem finding while at the same time retaining the thought of going beyond the question itself, FG1 came out with an agreement that involved adding “ideas” in the second half of the definition to allow for other discoveries and removing the phrase “the discovery of the purpose to be communicated” from the definition of problem finding. That is, the definition of problem finding was revised as “discovering problems or ideas in society”.

(2) Difference between these Components in Creativity and in Other Capabilities

A new comment on the revised definition concerned the difference between problem finding within the construct of creativity, relative to other capabilities.

> Could we add like “creative problem finding”, with adding “differently” at the end of the sentence to enhance the difference between problem finding in the normal way of thinking and in terms of creativity?

This query from Participant P03_FG1 can be extended to the definition of other components as a common issue because they didn’t highlight the characteristics under the construct of creativity. A qualified definition of each component in creativity should provide an image of what creativity looks like with respect to this component. The redrafted definitions for each component, which were then presented to the next focus group for verification, are presented in the following paragraphs.

Generally, problem finding in creativity requires an individual to be able to "think outside the box" to find a novel and appropriate problem that the individual student has never asked before. Novelty could emerge from a unique point of view by associating an individual's experience and knowledge, or through an analysis of a scenario that is different from what they did before. To highlight the extent of novelty in students, problem finding was redefined as "discovering novel and appropriate ideas or problems that the student has never noticed before".

Similarly, the component of resource expansion needed to be revised to fit with the construct of creativity. In this component, the student uses varying strategies to find feasible resources or information, such as experience and multidisciplinary knowledge, to generate creative achievement. However, only those not normally associated with the context are viewed as creative resources. They might come with various skills, such as analogy and metaphor (Finke et al., 1996; Sun et al., 2019a; Ward, 2007), or from a different perspective in analysing stimuli. Based on the above discussion, resource expansion in the construct of creativity is defined as **"**developing disparate appropriate resources that are novel to the individual student".

In resource expansion, students use their experience and knowledge to search for as many methods as possible and then evaluate them to find the most novel and appropriate way to combine available resources. Only unfamiliar approaches students had never thought of before can be considered creative combinations. Thus, combination searching for the purposes of this thesis is defined as **"**synthesising selected resources by using creative combinations that are novel to the individual

student and appropriate for the contextual purpose".

Due to finite quanta of time and energy, students should concentrate on the most creative rather than all the responses necessary to produce a creative achievement at the end of a process of generation, which indicates that they should know what it means to be creative in a way that facilitates the result. Students with evaluation skills can select the best idea worth pursuing after considering various possibilities and criteria. The "best" means the most novel and appropriate for the contextual purpose. The meaning of evaluation in this context is "the capability to identify novel responses that an individual student has never proposed before and defend the novelty of the response through sustained argument".

5.3.3 Conclusion regarding the Components of Creativity

The above results were shared with all the participants in FG1 to seek their feedback. They found a significant improvement in the revised definitions of these components. They were unanimously of the view that the current definitions of the components provided more specific and clearer meanings, which enhanced the likelihood that their measurement would more closely align with the construct of creativity (increased validity). As such, the definitions of these components were accepted by the participants in FG1. The definitions are provided in Table 5.1 below.

Table 5. 1 *Revised Definition of the Components of Creativity*

Component	Definition
Problem finding	Discovering novel and appropriate ideas or problems that the student has never noticed before.
Resource expansion	Developing disparate appropriate resources that are novel to the individual student.
Combination searching	Synthesising selected resources by using creative combinations that are novel to the individual student and appropriate for the contextual purpose.
Evaluation	Identifying novel responses that an individual student has never proposed before, and defending the novelty of the response through sustained argument.
Outcome	Producing novel and appropriate ideas, solutions, and products.

5.4 Reviewing the Draft Performance-Level Descriptors of Creativity

This section discusses the extent to which the draft performance-level descriptors provide an appropriate and explicit description of growth in each of the components in creativity. It first analyses the comments of the participants in FG1 on the performance-level descriptors that make explicit the performance standard of each component and provides arguments for retaining or modifying the descriptors to better describe growth. The integration of these amendments constitutes a new version of the generic rubric for measuring creativity.

5.4.1 Problem Finding

Three issues emerged from the qualitative content analysis of the discussion

carried out during this part of the focus group for “ask common problems” and “category” as well as taxonomic verbs focused on learning. The first two address the meaning of the descriptors at each level, while the remaining issue refers to the discussion of the extent to which these descriptors can reveal a learning continuum linked to the construct of creativity through the components.

(1) “Ask common problems within one category”

The discussion started with Participant P03_FG1’s admitting to some confusion, by saying:

> As a teacher, I can’t understand what you’re expecting me to think about the one “notice and ask common problems”, and the one “ask common problems among similar categories”.

Participant P05_FG1 also commented:

> … those categories don’t really mark the growth. I can’t say that there is a generation of growth or deep learning being recognised in the rubric.

Another pertinent issue that affected participants’ understanding was the ambiguity of using “common” or “uncommon” in the descriptors. While some participants suggested replacing these with other contextual words, such as “familiar and unfamiliar” or “typical and atypical”, Participant P05_FG1 expressed her view as follows:

> … that can be tricky because what defines “uncommon”? It might be common to me, but it might be uncommon for someone else. You need to support it with some contexts, and that’s what we are talking about …

> contextualise understandings or contextualise problems so that expands it into taking it from the learners' space and letting them see these problems in larger contexts so that they have that sense of how to expand into places outside of this …

Since the current descriptors can only provide the learning outcome achieved by the student at each level, Participant P01_FG1 suggested adding an explanation of "how" the student achieved the underlying cognitive thinking so that the terms "category" and "common" could be explained in a more cognitive context. FG1 argued that descriptors incorporating contextual explanations could be more easily used to make accurate judgements when teachers reference students' responses to the different performance levels of the performance standard.

This issue was addressed by recalling the operation of the three creative characteristics (fluency, flexibility, originality) using different weights to indicate different levels of complexity, thus showing growth in the component of problem finding (see Section 3.3.2.1). The "category" could be understood as "perspective" in the component of problem finding. It indicates that if the discovered problems to a task reflect the student looking at the scenario from a single point of view using only fluent thinking in creativity, then these problems will be described with reference to the Level 1 performance descriptor. These problems, categorised to analyse the task from a single perspective, are generally indicative of the initial thoughts of every student and reflect low cognitive complexity embodied in fluent thinking. To demonstrate growth along the learning continuum, students need to analyse the task

from different perspectives using flexible thinking. Students who can identify common or sophisticated problems that demonstrate their flexible thinking in creativity will be located at the middle level of performance in the component of problem finding. The higher level of performance descriptor is associated with the characteristics of originality, referring to students who can analyse the task and scenario through a broader lens and think outside the usual boundaries to discover novel and appropriate problems.

The above explanation was fully supported by the participants. They argued that incorporating this explanation into different levels of performance descriptors would provide the necessary contextual information on cognitive complexity and thus support the growth of creativity along the developmental continuum. As Participant P04_FG1 commented:

> … it makes much more sense to me when you [the researcher] explained. I was a bit lost in time to think what to do, and I think your explanation needs to put [it] there.

Therefore, the performance standard was revised by incorporating the above explanation into each level of performance descriptors. The performance standard is further elaborated at the end of this problem-finding section.

(2) "Category"

The term "category" or "categories" was used frequently in the performance-level descriptors in all the components. Multiple interpretations of the word "category" led to difficulties in conveying a clear and consistent understanding to

teachers, Participant P07_FG1 proposed the word “subject” to be used instead, to provide a more precise meaning of “category” in the school setting. However, other participants had different views on this issue, Participant P02_FG1 stated:

> This word “category” was troublesome for me, but as a generic rubric, it can include a lot of things, which means it can allow teachers actually to use it in their own specific subjects or disciplines …

And Participant P06_FG1 stated:

> When we talk about “categories”, we do tend to talk about grouping ideas and things of meaning. They might be of meaning, of similarity, of extent … that tells how distant they are … I think “categories” is appropriate.

These comments are consistent with the original intent of the descriptor design, which was to make the performance standards generic enough to be adapted to different contexts (e.g., subjects and scenarios). This was agreed by the participants in the focus group, and the term “category” has been retained in the performance standard.

(3) Taxonomic Verbs Focused on Learning

The participants critically reviewed the taxonomic verbs associated with learning used at each level of performance descriptor, which involved “asking” in Levels 1 and 2, and “generating” in Levels 3 and 4. Overall, FG1 advocated using different verbs from the Revised Bloom’s Taxonomy to indicate cognitive depth and learning development.

As the verb “asking” had been used in Level 1 to describe the basic requirement of problem finding in creativity, the participants did not agree with using the same word in Level 2. The same argument was given in relation to the term “generating”. While the word “applying” was nominated by Participant P04_FG1 to replace the repetition of “ask” in Level 2, Participant P01_FG1 considered it inappropriate because “you are not applying anything, you are asking them to explore something” and suggested using “explore” instead.

In terms of “generating”, Participant P08_FG1 proposed replacing “generate” in level 3 with “construct” because:

> As “generate” is a big word that includes higher order thinking, maybe you need to get a middle thinking verb fairly in Bloom’s Taxonomy … what about “constructing” in the third [level], I think that “construct” might be better in between the “explore” and the “generate”.

The suggested replacements above received more support than the initial verbs in FG1. According to Participant P07_FG1, “They show a nicer progression of deeper engagement and make deeper levels of inquiry …” Considering that the performance-level descriptors would be modified based on this focus group discussion and the suggested verbs might not fit the new context, Participant P05_FG1 reiterated the focus of the comments:

> … maybe there is another replacement or another synonym you can use for “asking” and “generating’ by searching Bloom’s Taxonomy. What we are trying to say is to use different cognitive verbs to show that learning

progression …

Based on this feedback, the learning taxonomic verbs at each level were critically reviewed by the researcher after supplementing the performance-level descriptors to help provide a clear distinction between the levels and show a general alignment with Bloom's Taxonomy to reflect the learning progression for the components. For example, the verb "ask" initially used in the descriptor of Level 1 performance was changed to "recognise" to align with the remembering level in the Revised Bloom's Taxonomy that describes the cognitive level of retrieving relevant knowledge from long-term memory. The same verb, "ask", used in the Level 2 performance-level descriptors continued to be used to indicate the understanding level, where meaning of ideas or situations is constructed by explaining, exemplifying, categorising, summarising, inferring, comparing and interpreting. The verb "develop" represents the applying level (i.e., develops a procedure by applying knowledge, skills, and techniques to solve problems in each situation). It was used in place of the original verb, "generate", in the descriptor of Level 3 performance as "generate" refers to the highest creating level in the Revised Bloom's Taxonomy. "Generate" could be retained in the Level 4 performance descriptor, but it was changed to "discover" to align with the verb used in the definition of problem finding, while keeping the cognitive level consistent with Bloom's highest creating level.

According to the feedback from FG1, though the initial descriptors in the performance standard of problem finding were considered to show differences across levels, contextual interpretations needed to be supplemented in the descriptions to

clarify their meaning. To enhance the extent of learning development in the performance descriptors, the taxonomic verbs were reviewed to unfold an explicit progression of cognitive thinking in creativity. Table 5.2 shows the amendment to the performance standard for problem finding, with the revised performance-level descriptions on the right.

Table 5. 2 *Amendment to the Performance Standard: Problem Finding*[1].

Level	*Original performance-level descriptor*	*Revised performance-level descriptor*
4	Generate a previously unknown, unexpected problem.	Analyses scenarios through a broader lens, thinks outside the usual boundaries and *discovers* novel and appropriate ideas or problems that few students have noticed before.
3	Generate uncommon problems in similar categories or ask about common problems in remote categories.	Analyses scenarios from dissimilar perspectives and *develops* sophisticated or unusual ideas or problems that are unfamiliar to peers.
2	Ask about common problems among similar categories or from similar perspectives.	Observes scenarios from similar perspectives and *asks* common ideas or problems that most students can notice.
1	Notice problems and ask about common problems within one category or from one perspective.	Looks at scenarios from one perspective and *recognises* superficial ideas or problems that every student can notice.

[1] Definition of problem finding: discovering novel and appropriate ideas or problems that the student has never noticed before

5.4.2 Resource Expansion

According to Participant P02_FG1, the prominent advantage of the draft performance standards of resource expansion and other components of creativity is that it promotes interdisciplinary learning in school education.

> I think [the use of] "discipline or category" is such a great thing, and it is closely related to current teaching issues. When teaching our subjects, the problem is that we are just standing in our own category. There is not much opportunity to cross over to other subjects, and students never have the chance to synthesise all the knowledge, trying to apply the knowledge to solve the problem and create a new idea and new product …

Given that these performance-level descriptors show appropriate learning growth in the component of resource expansion, the participants in FG1 emphasise the appropriateness of words used and the meaning of each descriptor. They focused on the term "appropriate" and the statement "connect usual resources in remote disciplines or categories" in the Level 3 performance descriptor.

(1) "Appropriate"

"Appropriate" and "novel" are two representatives of creative traits in the construct of creativity. The frequent use of the word "appropriate" in the description has caused some controversy in the focus group discussion. Participant P06_FG1 expected a more specific term to replace "appropriate" because "it is a bit ambiguous … how can we have an agreed term for the 'appropriateness'?" While

some participants advised specific replacements if necessary, such as “usefulness” and “related’, Participant P01_FG1 highlighted the domain generality of this rubric as follows:

> But this is a generic rubric, so it is specific to the teachers when they use it. You don’t need it so specific to “appropriate”, and it’s a good ambiguous word for this instead of using “usefulness” because you’re putting on a value to it. That’s a whole different arena of thought … teachers will sit there and go, novel and appropriate to the context of climate change, so you can’t go off and talk about something that’s not appropriate or unrelated to …

This argument was supported by Participant P05_FG1, who reviewed the descriptors through an arts lens, saying:

> I have a problem with the word “related”. The urge for creativity is about [being] creative. We don’t want to be related necessarily; we want to be novel, kind of out-of-the-box ideas, concepts, actions, or achievements related to the problem. I am now starting to wonder if a creative response is always responding to a problem, maybe it’s just a creative arts thing, an exploration of the other concept or an idea, or it is building a narrative around certain things …

To be generic, the creativity rubric should be appropriately broad for teachers to contextualise to different disciplines or settings when implementing it in their teaching. The general word “appropriate” was retained in the rubric as the above

comments were well supported in FG1.

(2) "Connect usual resources in remote disciplines or categories"

To begin with, FG1 considered the performance descriptor at Level 3 problematic and confusing and responded by suggesting a few synonyms for the word "remote" and antonyms for "usual". However, these modifications tended to completely change the meaning of the descriptor. The performance descriptor at Level 3 was originally developed to make two situations of learning performance explicit. An uncommon resource may be one that is not frequently used in similar categories, or it may be one that is common in other categories but transferred for use in this scenario. For example, to measure atmospheric pressure, students at Level 3 could perform the "Upside Down Glass of Water Experiment" in practice, which is common knowledge in physics instruction, but seldom can students think of implementing it. Another option could see students using injection syringes as one of the measurement resources, which is common in health and medicine but uncommon in high school physics.

After discussing what the descriptor meant, FG1 participants believed it was the statement, other than the wording, that needed to be rephrased to distinguish the meaning between these two alternatives, as Participant P05_FG1 commented:

> After listening to your [the researcher's] explanation, you need to rephrase the beginning of a sentence to differentiate it from the end of the sentence. Because I don't see enough difference between the first part of that sentence and the second part of that sentence to put an "or" in it.

An illuminating comment from Participant 04_FG1 resonated with the group, who questioned whether there had to be two specific parts in the description to clarify where the "unfamiliar" resources were coming from, as shown below:

> We can look at different ways of using the words to allow students and teachers to use these descriptions in their own contexts. We don't have to be specific in two parts … also, in my opinion, students who are able to transfer and apply knowledge or resource across different categories are creative enough to be located at the highest level …

Consistent with the teaching experience of the participants, this comment produced a satisfactory solution that retained and modified only the first part of the statement as the performance descriptor in Level 3 and placed the other half in Level 4 to explain the extent of novelty. With this amendment, the descriptors were demonstrating a more explicit meaning of student performance across different levels and effectively describe a meaningful learning progression in the component of resource expansion.

The original and revised performance-level descriptors in the component of resource expansion are shown below in Table 5.3. As can be seen, the basis of the amendment goes beyond the feedback for the component of resource expansion described above. This is in response to the feedback on using the word "common" that appeared in the problem finding and in the consideration that providing additional contextual information can better clarify the explicit meaning in the descriptors.

Table 5. 3 *Amendment to the Performance Standard: Resource Expansion*[2].

Level	*Original performance-level descriptor*	*Revised performance-level descriptor*
4	Develop and assemble disparate resources that are novel and appropriate to the context.	Applies a variety of skills and strategies to develop disparate appropriate resources across dissimilar disciplines or categories, that are novel to the individual student.
3	Associate unfamiliar resources from similar categories or connect usual resources in remote categories.	Uses basic skills and strategies to search and select unusual and appropriate resources across similar disciplines or categories, that are unfamiliar to peers.
2	Collect common resources across similar categories or disciplines.	Associates similar situations and collects typical resources across similar disciplines or categories, that most students know.
1	List common resources within one discipline or category.	Recalls identical situations and lists common resources within one discipline or category, that every student knows.

5.4.3 Combination Searching

The performance standard for combination searching was considered to show clear differences across performance levels. The following feedback was received in discussion within FG1.

I would like to apply some verbs in Bloom's Taxonomy as much as

[2] Definition of resource expansion: developing disparate appropriate resources that are novel to the individual student.

possible. You can use the verb “apply” here in the second, so the first one is “follow” and the second is “apply”.

This first comment, from Participant P08_FG1 relates to the repeated use of “follows” in the performance descriptor at Level 2. Another comment was made by Participant P01_FG1, who considered the phrase “follow methods previously existing” backwards and suggested using simple and concrete phrases to replace it, if possible, as it was used in various levels of performance descriptors.

The solution to this suggestion was to contextualise the actual meaning of this phrase in the descriptors by rethinking students’ performance in combination searching at each level. The meaning of “previously existing methods” refers to the same methods as intended to describe students at the basic level. The contextual meaning at Level 2 includes analogy learning, in which students are believed to relate the scenario to various similar situations and apply typical methods without modification to combine resources. At Level 3, students can modify and adjust these methods to fit the scenario better. There is no need to use this phrase at the highest level because novel approaches come from integrating different methods across different disciplines or categories. The above analysis, together with the recommendation of using the verb “apply” in the second level, leads to the revised performance standard of combination searching in Table 5.4.

Table 5. 4 *Amendment to the Performance Standard: Combination Searching*[3].

Level	*Original performance-level descriptor*	*Revised performance-level descriptor*
4	Create a novel and appropriate way that draws on previously existing strategies to synthesise resources.	Creates disparate appropriate methods by integrating various approaches across dissimilar disciplines or categories to synthesise resources, that are novel to the individual student.
3	Implement organised ways that use a combination of previously existing strategies to synthesise resources.	Develops unusual and appropriate methods by adjusting similar approaches across similar disciplines or categories to synthesise resources, that are unfamiliar to peers.
2	Follow ways previously existing within categories or domains in similar situations to combine resources.	Applies typical methods that students have known before in similar situations to synthesise resources.
1	Follow ways previously existing within a category or domain in an identical situation to combine resources.	Follows the same methods that students have known before in identical situations to synthesise resources.

[3] Definition of combination searching: synthesising selected resources by using creative combinations that are novel to the individual student and appropriate for the contextual purpose.

5.4.4 Evaluation

Most of the participants in FG1 believed that the descriptors in the draft performance standard for the evaluation component provided a clear differentiation between each performance level. However, Participant P05_FG1 considered the term "relatively" too broad and therefore, "it could prevent teachers from professionally referencing students' performance to Levels 2 and 3 descriptors". In addition, the point was made that relying solely on a student's ability to provide logical justifications for their choices does not adequately distinguish their level of creativity. The novelty of their chosen answers needs to be taken into consideration.

One suggestion from FG1 was to replace "relatively" with the following hierarchical progression: "very", "significantly", "extremely" and "strongly". However, these adjectives are more intense and correspond to the highest levels on a continuum characterised by increasing levels of cognitive depth. Another more feasible solution, as proposed by Participant P05_FG1, was to "clarify a contextually appropriate response to that level". Instead of using loaded words that requires teachers to make judgements based on their "gut feelings", the wording in the descriptors should be clear and measurable to describe what students know and can do.

The underlying philosophy for addressing this issue is the same as the previous modification of the performance-level descriptors for the other components (e.g., "common" in problem finding, "previously existing methods" in combination searching), which is to provide more contextual cognitive information to the generic

words. To develop creativity in learning, students need to know what creativity is and have a basic idea of their own performance. Students who lack knowledge of creativity may choose a sophisticated response when measuring the “evaluation” component because it would be the best response to get a high score on a standardised test. While sophisticated responses imply higher order thinking such as analysing, evaluating and creating, creative answers go beyond the realm of sophistication and include the characteristics of being unusual or novel to learners themselves.

Table 5.5 shows the original and revised performance standard for the evaluation component. The extent of being “relatively novel” in the original performance descriptors in Levels 2 and 3 was clarified by the extent to which students were able to provide supporting arguments for their choices. This amendment was believed to make clearer distinctions between different levels of performance descriptors to show the growth in what students should know and can do at each level of performance in the component of evaluation. This distinction makes it more likely that teachers would be able to locate students more accurately and reliably on the performance standard by referencing their responses to the performance-level descriptors.

Table 5. 5 *Amendment to the Performance Standard: Evaluation*[4].

Level	*Original performance-level descriptor*	*Revised performance-level descriptor*
4	Choose the most novel and appropriate response with a persuasive explanation	Chooses a novel and appropriate response that has not generally been proposed before, and provides arguments to support its novelty.
3	Choose a relatively novel and appropriate response with a reasonable explanation.	Chooses an unusual and appropriate response by comparing it with peers' answers, however, the argument defending its novelty is not well sustained.
2	Choose a relatively novel and appropriate response without clear reasons.	Chooses a sophisticated response that *most* students will also identify as their solution.
1	Choose an appropriate response to the context with conventional reasons.	Chooses a common and appropriate response that is known by every student.

5.4.5 Conclusion regarding the Draft Generic Measurement Rubric for Creativity

No amendment was made for the descriptors in the performance standard of the outcome because they were evaluated as being well-described in meaning conveyance and hierarchical differentiation. The validation of the measurement rubric

[4] Definition of evaluation: identifying novel responses that an individual student has never proposed before, and defends the novelty of the response through sustained argument.

primarily revolves around two themes. The first is the extent to which the performance descriptors provide a true and faithful delineation of what it means to grow in relation to each of the components. The second theme is the extent to which there is clear differentiation and equidistance between the performance-level descriptors comprising the performance standard for each of the components. Much effort was expended on addressing the first theme of providing appropriate and explicit descriptors as it is fundamental to examining further whether they show a learning progression that enables an accurate judgement to be made by teachers. Regarding the second theme, the participants argued that it would be too difficult to distinguish the growth of each of the components of creativity if there were five levels of performance in the measurement rubric. The distance between the second and third levels would be greater than the distance between the first and second levels if there were only three performance levels. They agreed that the revised measurement rubric, categorising the components of creativity into four performance levels, provides equidistant distances between the different performance levels.

The amendment largely involved specifying the meanings of ambiguous words or phrases at each performance-level descriptor. After the meeting, the revised rubric was shared with FG1 to gain further feedback. The participants were satisfied with the new version of the measurement rubric because they felt as though there had been a positive influx of contextual information and they felt the more detail the descriptor provides, the easier and clearer it is to understand. Participant P01_FG1 supported made the following point:

> This explicit generic rubric gives me enough information and a broad range to look at students' creativity, then I can adjust and narrow it down to a specific context.

A revised generic measurement rubric for measuring creativity in students is shown in Table 5.6.

Table 5. 6 *Revised Generic Measurement Rubric for Creativity.*

Definition of creativity:

Creativity is the capability to generate novel and appropriate outcomes by employing problem finding, resource expansion, combination searching and evaluation.

Components:

- Problem finding: discovering novel and appropriate ideas or problems that the student has never noticed before.
- Resource expansion: developing disparate appropriate resources that are novel to the individual student.
- Combination searching: synthesising selected resources by using creative combinations that are novel to the individual student and appropriate for the contextual purpose.
- Evaluation: identifying novel responses that an individual student has never proposed before, and defending the novelty of the response through sustained argument.
- Outcome: producing novel and appropriate ideas, solutions, and products.

Component	Level 0	Level 1	Level 2	Level 3	Level 4
Problem finding		Looks at scenarios from one perspective and raises superficial ideas or problems that every student can notice.	Observes scenarios from similar perspectives and recognises common ideas or problems that most students can notice.	Analyses scenarios from dissimilar perspectives and identifies sophisticated or unusual ideas or problems that are unfamiliar to peers.	Analyses scenarios through a broader context, thinks outside the usual boundaries and discovers novel and appropriate ideas or problems that few students have noticed before.

Resource expansion	Recalls identical situations and lists common resources within one discipline or category, that every student knows.	Associates similar situations and collects typical resources across similar disciplines or categories, that most students know.	Uses basic skills and strategies to search and select unusual and appropriate resources across similar disciplines or categories, that are unfamiliar to peers.	Applies a variety of skills and strategies to develop disparate appropriate resources across dissimilar disciplines or categories, that are novel to the individual student.
Combination searching	Follows the same methods that students have known before in identical situations to synthesise resources.	Applies typical methods that students have known before in similar situations to synthesise resources.	Develops unusual and appropriate methods by adjusting similar approaches across similar disciplines or categories to synthesise resources, that are unfamiliar to peers.	Creates disparate appropriate methods by integrating various approaches across dissimilar disciplines or categories to synthesise resources, that are novel to the individual student.
Evaluation	Chooses a common and appropriate response that is known by every student.	Chooses a sophisticated response that most students will also identify as their solution.	Chooses an unusual and appropriate response by comparing it with peers' answers, however the argument defending its novelty is not well sustained.	Chooses a novel and appropriate response that has not generally been proposed before, and provides arguments to support its novelty.

Outcome		Copies or restates an idea, solution or product from existing material that is mostly appropriate and contextualised.	Develops an idea, solution or product that is a conventional combination of something existing, and/or is mostly appropriate and contextualised.	Generates an idea, solution or product that shows novel aspects and is appropriate and contextualised.	Creates an idea, solution or product that is entirely novel, highly appropriate, and contextualised.

5.5 Chapter Summary

In FG1, participants from a variety of disciplinary backgrounds provided feedback and evidence to support the validation of the draft construct of creativity and its measurement rubric according to their teaching or research experience in high school contexts. Necessary and appropriate amendments were made to the definition of creativity, its components, and the performance standards, to describe the construct of creativity and its growth to teachers in a clearer manner that was easier to understand. These amendments, which were highly supported by the participants in FG1, enhance the validity of the generic measurement rubric for creativity, making it more effective for use in classroom assessment.

The result of the FG2 discussion is presented in the next chapter. The comments of the FG2 participants are categorised and quoted in the same way as the comments of the FG1 participants in this chapter, for the purpose of validating the generic measurement rubric for creativity to be used in the context of science.

Chapter 6 Validation Analysis: Science Focus Group

6.1 Introduction

This chapter presents the results from Focus Group 2 (FG2) as part of the evidential case supporting the validation of the measurement rubric and its appropriateness for use in science teaching. These results further answer the second research question, from the discipline of science, of the extent to which a generic measurement rubric for creativity enable the measurement of growth in creativity across disciplines. The participants in FG2 were six educators with a background in high school science teaching or research. They critically reviewed the extent to which the generic measurement rubric for creativity, as revised with validity evidence provided by FG1, could be used in the domain-specific context of science.

The results are presented in the same manner as those from FG1 in Chapter 5. It consists of three sections based on the predetermined structure of the discussion questions (see Appendix E): the draft definition of creativity, the defined components and their performance-level descriptors, which constitute the performance standards.

6.2 Reviewing the Draft Definition of Creativity

FG1 produced the draft definition of creativity as “the capability to generate novel and appropriate outcomes by employing problem finding, resource expansion, combination searching, and evaluation”. The participants in FG2 began by reviewing both the structure and meaning of the FG1 definition to determine if the draft

definition is appropriate for use in the science domain.

6.2.1 Structure and Meaning of the Draft Definition

The participants in FG2 considered the structure and the meaning of the draft definition to be well described and consistent with their understanding of creativity. Participant P01_FG2 stated that the creative potential and its described outcome could lead to a holistic investigation of student creativity:

> I like it a lot. I think it has both ends of the scale from what is capable and what should happen as the outcome. I think it's perfect, and I wouldn't change it.

Similarly, Participant P03_FG2 acknowledged the value of including process-related components in the definition from a practical pedagogical perspective:

> I think having the process and the outcome together [in the definition] is important. In science, a lot of people think that the end product matters, and it doesn't matter how kids get the right answer in the exam, whether you've made [them] memorise three lessons or whatever …

While agreeing with the structure of this definition, Participant P02_FG2 shared another definition of creativity that is popular in social science as a reference and asked about the purpose of this draft definition and its accompanying rubric:

> I think this definition is good. Another definition I like in creative thinking is … a habit of mind. It's not so much that people have to be creative thinkers on a particular task, but that they do it routinely … the

> question is whether you want the rubric to apply to an individual, how they've been over a term, or do you want it to apply to an individual task. Because if it is over a term, then you say, they habitually or frequently do these things. Then the development and progression are wrong, maybe occasionally doing it through to having it highly frequent …

The idea of connecting creative thinking on a broad day-to-day basis relates to the "Little-c" level of creativity according to the Four C Model of Creativity, where the novelty and appropriateness of its creation are of value to others (Kaufman & Beghetto, 2009). However, this thesis focuses on creativity in everyday learning – "Mini-c" – one developmental step before "Little-c". Students at the "Mini-c" level must be encouraged to think creatively and cultivate their creativity in everyday learning. In this process, teachers should be able to use the measurement rubric to evaluate how well the student has performed and to design tailored instruction for developing student creativity after getting feedback from the performance of the students. The measurement rubric captures a learning progression for student creativity that enables school teachers to better understand and monitor student development of creativity by contextualising and embedding the rubric in their teaching contexts. Students who receive scaffolding support after the assessment activities would be able to develop their creativity from low to high along the measurement rubric for creativity and thus, move forward to the next developmental step – "Little-c", which describes everyday creativity. This step of day-to-day creativity, as suggested by Participant P02_FG2, could be defined in terms of how

often creative ideas are generated. In addition, the participants were generally of the view that the frequency of using creative thinking should not be a component of growth in creativity.

The above consideration was accepted by Participant P02_FG2, who believed that the draft definition provided by the researcher could be better aligned with the purpose of this study. Participant P05_FG2 also supported using the draft definition developed by FG1 by providing an example of using this definition to inform the design of an instructional task in their classroom:

> I think teachers are supposed to use it [the draft definition] for a task … the teacher uses this and says, "I'm going to teach my students volcanoes, and I want them to design a volcano, so I would use your rubric to measure the student's ability to create that volcano for that one task" … This is much more useful for teaching and learning.

Participant P06_FG2 also prioritised the draft definition over others because it is aligned with teaching practice and highlights the vital characteristics of creativity:

> I think it [the draft definition] is very good. It covers a lot of things, and having novel and appropriate together is important. Because you can't just have "novel" that gets a new idea, and it's not appropriate to the context. I think the words "novel' and "appropriate" are good, and the definition is very clear.

The draft definition of creativity remains unchanged at this stage as the participants unanimously agreed that its structure and meaning are appropriate for use

in the science domain. There was some feedback about the similarity of the definition to students' general learning, which is analysed and presented in the next section.

6.2.2 Relationship between Creativity and Learning in General

> The definition is not much different from the learning in general. The word [that] really makes the difference is "novel" there. Apart from that, you can describe any learning … even normal learning, if it is new learning … [it] is novel anyway …

The focus group discussion revolved around the above comment from Participant P02_FG2. On the one hand, learning is always new to students, as suggested by Participant P04_FG2:

> … it is true to the term "novel" because everything is novel to a student if they haven't done it before… In the science context, a lot of what students do [is] novel. They are creating something, being problem solvers, or knowing how to clean up an oil spill … science experiments and activities are novel to students until they do them or hear about them.

However, as knowledge in domains has been determined by the curricula and associated syllabi, the focus of creativity education is not on knowledge itself but on the way the material is taught and on the tasks that teachers use to promote and assess creativity, according to Participant P03_FG2:

> I disagree with others. There is a world of difference between giving the answer versus letting kids have a goal and seeing what they come up

> with. Imagine the scene of an oil spill in Sydney, and teachers tell kids how to clean it up, versus here is a video of me cleaning up an oil spill and asking kids to write down their ideas before the class ends. This example may not be about creative thinking, but it shows a different way of learning …

This view was also supported by Participant P06_FG2:

> It could be novel to students in terms of the content but, for creativity, it depends on the questions you ask. If you ask kids something that can be easily found in the textbook, then they'll just show you the answer that everyone knows …

This discussion section reflects a common issue of fostering creativity in general teaching and learning. Though school teachers have recognised the importance of creativity for students' life-long development, they find it difficult to design and implement appropriate instruction in relation to creativity in the teaching of their subject (Cropley et al., 2019; Harris & Bruin, 2017). While designing specific creative projects can boost students' motivation and knowledge of creativity, creativity is essentially the same as any other capability that can be developed in everyday teaching (Perkins & Perkins, 2009; Weisberg, 2006). Participants P03_FG2 and Participant P06_FG2 in their examples above, make the point that, in their opinion, creativity or creative thinking is more about how to apply different knowledge and skills to new situations. Therefore, creativity can be stimulated and investigated by designing appropriate assignments or questions (or tasks in a broad

sense) or by adjusting and implementing appropriate pedagogical approaches. In other words, they argued that it is unnecessary and impractical to distinguish creativity from general learning process in the definition.

In general, in the teaching and learning process, teachers have unlimited opportunities to foster and measure student creativity by adapting each piece of learning material appropriately. The same learning materials can be organised in a traditional way with the purpose of firm and quick retention of knowledge or in another open-ended, challenging way to allow students to come up with novel, appropriate responses. Thus, the draft definition of "creativity" remains unchanged to indicate that creativity can be cultivated and measured during the general process of learning, with the term "novel" being used to highlight the characteristics of creativity.

6.2.3 Conclusion regarding the Definition of Creativity

No amendments were made to the definition of creativity, as FG2 participants largely agreed that it provides a clear description of student creativity and is appropriate for measuring creativity in the context of high school science. The group discussion then moved on to a review of the appropriateness and practicality of its components and their definitions, which is analysed in the next section.

6.3 Reviewing the Components of Creativity

The participants in FG2 then examined the alignment between the definition

and its components. Comments such as “all these components make up the definition of creativity, which shows a clear consistency” and “these definitions highlight the characteristics of creativity very well” from Participants P06_FG2 and P05_FG2, respectively, indicated an agreement that the components are aligned with the definition of creativity.

While the participants valued these components for constructing student creativity, they were inclined to consider the extent to which the components can be measured in their teaching practices. For example, Participant P01_FG2 expressed their preference for the components of resource expansion and problem finding:

> I quite like that [the components and their definition]. Like resource expansion, a lot of kids now are extremely good at searching, but surely anything except what Wikipedia or the first three Google hits they get … I don’t think there is anything bad about them [the components] as far as identifying what creativity means …
>
> In science, I think that some of the better kids do have a different way of thinking about problems, and I think that “problem finding” is quite important, too … we teach some equations or all the chemicals or whatever, but the problem finding is a nice one, I like that probably the best of all of these. A lot of students, if they’re quite young, think probably there is nothing more to learn. In science, someone has done it before in some ways, but I haven’t. So, problem finding is a very good component to emphasise.

This steam of thought was supported by Participants P04_FG2 and P05_FG2, who said that “they [the components] are very doable in a science classroom” and “I can quickly relate them to my science class and design tasks to measure the skills of problem finding and resource expansion in creativity”. However, when the participants moved to the next three components: combination searching, evaluation and outcomes, there was a brief silence in the discussion, indicating that participants encountered challenges in connecting these components to science classes. Participant P05_FG2 stated:

> It could be difficult for teachers to measure creativity through the lens of combination searching because we usually combine resources in our own disciplinary ways, and that’s scientific but not creative … As a teacher, we should also open our minds to think about interdisciplinary learning and STEM-like projects in school to cultivate and measure students’ creativity … I think combination searching can be measured in these projects …

Participant P03_FG2 emphasised the value of the evaluation component because it reminds teachers of the process used by students to produce ideas or perform tasks, rather than focus on the responses themselves:

> It’s almost like you want new ideas from them, but how can you approve that? They could come up with a clever idea and build it from the ground up … you have to know that maybe it’s from the textbook, but they’ve never seen that textbook, so it’s new to them. Evaluation gives us a

> chance to ask students "how" they build up the idea or anything … It is important.

In terms of the outcome component, Participant P01_FG2 found it unclear even though it had been specified as "ideas, solutions and products". While the participants in FG2 acknowledged the diversity of creative outcomes, Participant P02_FG2 suggested more examples could help teachers better integrate the outcome process into their classrooms:

> In the outcome stuff, perhaps one of the reasons why that seems the most difficult component is that we could have a little bit more on what sort of outcome it is. Is it a communication outcome or an object outcome? Is it a product outcome with [an] aesthetics consideration … maybe some additional detail on that would help, even though you've got the definition to bring it through to the rubric.

A similar comment, indicating that it is difficult to understand and set appropriate tasks for the component of outcome, was first advanced in FG1. It indicates a common phenomenon that teachers are more acquainted with teaching and measuring something concrete and constant, such as well-structured disciplinary knowledge and step-by-step scientific inquiry skills. Including more examples of various creative outcomes to give more clarity to the definition effectively provides teachers with the opportunity to expand their thinking and design appropriate tasks. However, producing an exhaustive list of creative outcome examples or exemplars would be difficult, even in the science domain. Creative outcomes can take a variety

of forms, such as physical (e.g., robot cleaners), written (e.g., lab reports), or oral (e.g., conservation campaigns). Participant P06_FG2 said:

> I think you can put some examples along with the definition, but for me, using this general definition [of the outcome component] is good because it allows teachers to design tasks appropriate to their teaching.

In conclusion, although the FG2 participants considered all the components were essential to the definition of the construct of creativity used in this study, the feasibility of their assessment in science classes was a point of concern. The components of combination searching, evaluation and outcomes were considered more challenging in science teaching than problem finding and resource expansion because few of the existing teaching and learning activities are related to these components. However, participants were also of the view that it would not be impossible for teachers to think of relevant assessment tasks if they were open to new ways of thinking. They were also of the view that the measurement rubric and the explication of the components might be the stimulus for teachers to try some new types or modes of assessment that could provide students with the opportunity to produce evidence associated with the three challenging components. FG2 participants reached an agreement to keep all five components and their definitions as part of the construct of creativity. One amendment was made to the outcome component by adding several examples at different stages along the measurement rubric. These examples include written expressions such as lab reports and essays; oral expressions such as communication approaches and singing; and physical products such as

handicraft, painting and robot cleaners.

6.4 Reviewing the Performance-Level Descriptors of Creativity

Next, FG2 participants were asked to discuss the extent to which the measurement rubric provided clear and appropriate performance standards and the extent to which the rubric could be used to measure student creativity in science instruction. The themes that arose in the discussion include the significance of a measurement rubric for creativity, the extent to which the performance descriptors could be used in teaching practice, the normative nature of the current performance descriptors and the emphasis on independent effort in the performance descriptors for problem finding.

6.4.1 Significance of Having a Measurement Rubric for Creativity

The participants considered the school-level measurement rubric for creativity significant for several reasons. School teachers may be able to measure and mark students' creativity, but they lack the knowledge to understand the developmental continuum that describes the growth of creativity and clearly describes the difference between the performance-level descriptors (and hence the marks). Participant P01_FG2 stated:

> We all give our assessments three marks for creativity [based] on teachers' gut feeling. It would be nice to have a more rigorous understanding of what students ought to be aiming for in creativity …

Participant P04_FG2 added to this comment:

> That's true, like a teacher saying you got 3 out of 4 there, and the other person got 4 out of 4. This rubric is very helpful if the person who got 3 out of 4 can see or articulate the difference.

In the absence of professional support, school teachers rely primarily on their own "gut feelings" to identify student creativity and give grades for students' creative performance. Participant P02_FG2 continued:

It is not a bad thing … The teacher has to rely on "gut feelings" because none of the assessments actually have framed creativity as important, even though it is a national goal. That's what we're trying to do … until we start talking about and defining and maybe even assessing creative learning, we won't have the system-level support to do that.

A measurement rubric allows teachers to step away from their "gut feelings" and make professional judgements about student creativity by referencing student responses and submissions (evidence) to the performance descriptors. It also contains a clear structure for measuring creativity that provides teachers with guidelines for designing creative tasks. Participant P03_FG2 commented:

> We don't have criteria like the definition you've got there, nor did we have specific criteria to assess and mark students' submissions. In practice, we may not assess a particular aspect of creativity in detail but only provide an opportunity for students to demonstrate it.

This comment may partially respond to the phenomenon of previous

discussions about the appropriateness of using these components in science teaching. Problem finding has always been the major aspect of creativity education and assessment (Guilford, 1950, 1956; Kaufman & Beghetto, 2009; Sternberg, 1986). Teachers who are aware of the significance of creativity are familiar with the concept of problem finding and have even been able to design learning tasks related to it in their teaching activities. Therefore, the participants in FG2 were able to react positively to this component and see how it can be used. Resource expansion is also generally assessed when measuring creativity, meaning that participants (who were teachers of science) could relatively readily connect it to the classroom. There was a general belief among the participants that teachers will generally struggle to think and design tasks for those "unfamiliar" but important components, such as the three they had been discussing in detail.

The participants identified two strengths of the measurement rubric in fostering and assessing student creativity – providing performance descriptors of developmental growth in creativity to demonstrate the performance difference behind codes, and providing a clear structure for designing learning activities and tasks to assess creativity.

The next section presents the group's discussion about the extent to which this rubric can be disseminated and used in science classrooms.

6.4.2 Practicality of the Generic Measurement Rubric for Creativity

When the participants were asked to discuss the extent to which the

measurement rubric could be used in their classrooms, they raised several concerns about what they thought could hinder constructing assessments that enable the students to provide evidence of their location on the measurement rubric. One of the challenges relates to the extra effort required to understand each component and descriptor in the measurement rubric. As Participant P05_FG2 commented, “I’m not sure that teachers would automatically understand these components and their descriptors”. This idea was supported by Participant P02_FG2, who stated, “The meaning of each concept needs time to understand it, and teachers need support in doing this …”

Participant P01_FG2 raised their concern from the perspective of the current syllabus and provided a possible solution for applying the measurement rubric in the following comment:

> It’s very hard to squeeze it into the syllabus, certainly in my school. Using the rubric at first is difficult, and teachers may not know the exact meaning of each performance until we start to reference students’ responses to the descriptors.

This idea echoes the theory of “Learning-by-Doing” (Freire, 1982) and its extended method, “Professional Learning Communities” (DuFour & DuFour, 2013), which highlights the fundamental role of interacting with the appropriate associated real-world environment in learning new things. Since creativity has not been given attention in daily science education and is not widely included in mainstream educational tests, most teachers lack expertise in assessing creativity, and they also

don't have wide experience in using observation to recognise students' performance with respect to creativity in their classes. Hence, initially, teachers may not fully understand a range of foreign concepts and descriptors until they trial them in instructional activities such as project-based learning. In this process, teachers would potentially gain more professional knowledge about students' actual performance of creativity. By using exemplars to illustrate different levels of performance, teachers will be able to build up an operational image of what each performance descriptor means in terms of creativity in their context.

The attempt to apply this rubric in student research projects reminded Participant P06_FG2 of the difficulty of building a scoring rubric for a STEM project at their school. They found it difficult to adapt and apply this rubric to some of their school's current programs, saying:

> I think a rubric should be simplified like a chart. When teachers glance at it, they know where their students live and set the task for the project. This one may be the wording, or it has too many components that we would need more time to identify those areas … it wasn't like I could pick up a student's past [work] and I could relate [it] to this rubric immediately …

The expectation of having a well-prepared measurement rubric that could be used directly for every type of creative project was opposed by Participant P05_FG2 as follows:

> I think it's hard to have the same rubric for every type of situation and

> project, like the STEM project you mentioned. They are big projects and therefore evaluate not only creativity but also other indicators that may vary from case to case … It's just that everything certainly has to be read a little bit into it about how it applies.

Participant P04_FG2 highlighted the benefit of having a generic rubric for measuring creativity in terms of the divergence of science teaching:

> I think if a teacher is going to use this for whatever they're going to design a task for, keeping it generic would be easier than being highly specialised because you will need to manipulate it to fit the task that you're going to do. When I am looking at this, I could probably know where most of my students would be. Every science teacher will be different, especially in the years they teach and the subject they teach.

Other members of FG2 agreed with these two views, that having this generic rubric can be more easily adapted and applied to varying science classes even though school teachers would need to invest more effort in understanding and practising this measurement rubric. In other words, the participants considered it unnecessary to establish a domain-specific rubric for measuring creativity. The discussion made it clear that as the rubrics become less generic and more specific (domain-specific or task-specific) the workload associated with constructing such rubrics increases and this lessens the likelihood that teachers will use a measurement rubric to monitor student creativity and increases the likelihood of reverting to "gut feeling".

The third theme related to the normative nature of the standards statement is

discussed in the next section.

6.4.3 Normative Statement in the Descriptors

The final phrase, “that few students have noticed before”, in the Level 4 performance descriptor of problem finding, generated an intense discussion, especially when the participants identified similar statements that appeared in various other performance descriptors in the rubric. As pointed out by Participant P02_FG2, this statement indicated the idea of a peer comparison that ran counter to the initial desire to establish a standards-based rubric.

> It sounds different to the other students they [teachers] teach, which is [an] undefined quality, so we don’t know what age bracket they’re being referenced to or anything like that … You want standard-referenced measures to create a transparent understanding of the components and growths, but this growth is described with the normative angle.

Participant P01_FG2 misinterpreted this statement as, “This is novel compared to what the rest of the class did or what the same class did last year”. It is therefore apparent that this statement is misleading because the degree of creativity at the student level should for the individual student and not require a comparison with peers. Some participants reflected on this group reference considering the nature of creativity. For example, Participant P05_FG2 commented:

> Is it because when we talk about creativity, we’re talking about being different or being unique and having unique ideas? To do that, you do

> need to compare to other members of the group … that's a condition. One's thinking would go into it, analysis of the difference, the chemistry behind it … You can't determine whether something is novel if you don't compare it to what's been done already, you have to compare it to something. I could see why you bought those different levels in there, but I can understand why the other teachers were talking …

Participants P03_FG2 and P06_FG2 were also of the view that certain norm-referenced descriptions are needed when determining what is unexpected in relation to creativity. However, they also made the point that normative judgements are dependent on the teachers' experience with comparative student populations. When searching for better replacements for this "peer" phrase, such as "commonly unknown", "unexpected by students" or "generally unfamiliar to students", the group realised that these similar expressions are still making comparisons to a cohort of students and could not be completely divorced from normative references.

Another possible solution was suggested by Participant P05_FG2, who recommended simply deleting the phrase in the descriptor:

> … maybe you don't need to have it at every time, like you don't need to have it where you have synthesised because when you are synthesising something, you are creating something new essentially by putting things together, so maybe you don't need it there.

Other participants immediately supported this idea. Participant P02_FG2 said, "It [the normative phrase] is not necessary to be there so that your descriptors can

focus more on the cognitive part you have." Participant P01_FG2 also reviewed all the performance descriptors and recommended eliminating such phrases from most Level 3 and 4 descriptors to better align with the requirements of standards-based assessment.

The tense relationship between norm- and standards-based assessment can be reflected by considering the process of establishing a standard. When building up a standard for a construct that is unfamiliar to teachers or researchers, they will imagine what students know and can do based on their experience with a normative group of students and provide a description that characterises each level (e.g., "very good") of performance (Hammer, 2007; Hendry et al., 2012; Suskie, 2018). As a consequence of this discussion, most of the normative statements in Levels 3 and 4 were removed. In contrast, those for Levels 1 and 2 remained to strike a balance between pursuing clearer descriptors and mitigating the impact of normative information in the measurement rubric.

6.4.4 Emphasis on Independent Effort in the Descriptors of Problem Finding

> … do you want the kid to get the most marks if they've developed a sophisticated answer without relying on other people's ideas? If you want them to build a volcano, and they google, they may realise that you combine carbon vinegar and food dye, and it's a volcano. If this student came up with all of that by themselves, you know they noticed, they brought all these threads together, if all the development was the

> student's, then that's indicative of a high level, rather than just copying somebody else's good ideas.

Participant P03_FG2 added:

> When we do a depth study, we usually attach little class tasks to that, so we can put the kids on the spot and ask them to explain one feature of the task and their reasoning behind their approach, thinking trying to prevent kids from copying. If you're building an idea like the protein synthesis model in biology, which is a Higher School Certificate (HSC) topic in NSW, my teachers have gone to a lot of trouble to try and make an assessment task that allows for creativity. We don't want something out of a textbook. That's how we try to put them on the spot because if you've done it yourself, you'll know exactly what you're thinking.

Participant P03_FG2 further argued that during the high-pressure year for Higher School Certificate (HSC), the assessment on creativity should emphasise how students generate their ideas when learning objectives are confined to textbooks. As a teacher, she did not want the students to ask their parents or google for the answer and replicate that idea. The group participants unanimously agreed with this and suggested adding the term "independently" to the Level 3 performance descriptor. Participants P05_FG2 and P06_FG2 agreed with the introduction of the word "independently" to eliminate ideas copied from others. This was supported by Participant P02_FG2 because "we use that ['independently'] a lot in the university rubrics, meaning within the student's own thoughts and ideas without outside help".

Participant P04_FG2 expressed their concerns about using the term "independently" as it seemed to contradict their preferred collaborative approach to teaching and learning. However, Participant P02_FG2 helped clarify the meaning of "independently" as follows:

> … "independently" can also be those three kids over there. You can say google the answer, or they just went to the group and collected ideas. Maybe "collectively independent" or something to say they bounce ideas off each other. There is no opposition to collaboration. "Independently" means without outside information or ideas.

To embrace the possibility of collaborative learning to develop independent ideas, Participant P02_FG2 suggested using "student/s independently" to consider a group of collaborative activities. The participants fully accepted this modification, with Participant P04_FG2 being satisfied: "It gives teachers a choice. I can see how this works in my collaborative groups." When asked if the same changes should be made to other performance descriptors, the participants considered them well described and did not require further additions. Therefore, the third level's performance descriptor of the problem-finding component was amended to "Student(s) independently analyse scenarios from dissimilar perspectives and identify sophisticated or unusual ideas or problems".

6.4.5 Conclusion regarding the Generic Measurement Rubric for Creativity

FG2 discussion provided some clarification about the meaning of the various components and their performance standards. The resulting changes with the associated reasons for making these changes, form part of the evidential case supporting the validation of the measurement rubric, its appropriateness for use in science teaching and the appropriate meaning of the performance descriptors. Further, the group judged that the generic measurement rubric could be effectively used in science teaching.

The participants in FG2 also suggested several minor modifications related to the normative statements in the high-level performance descriptors and the emphasis on thinking independently with respect to the component of problem finding. To be specific, most normative statements from Levels 3 and 4 descriptors were removed to be more consistent with the philosophy and requirements of standards-referenced assessment. The phrase "student(s) independently" was added to the Level 3 descriptor for the problem-finding component, to highlight the independent effort of developing sophisticated ideas or problems within students' own thoughts. The new rubric was shared with FG1 and FG2 and finalised with their agreement. Table 6.1 shows the new version of the generic creativity rubric, with the changes in red, italic and strikethrough.

Several creative activities shared by the participants were documented and used as examples for discussion in the subsequent teacher development workshops (see Section 7.2.1). For example, Participant P06_FG2's STEM project of designing a

vacuum cleaner and building a model of the continent for plate tectonics from Participant P03_FG2 could be used to measure the outcome component; Participant P05_FG2's soil erosion activity of asking students to find different materials for investigation could be related to the assessment of the resource-expansion component.

Table 6. 1 *Updated Generic Measurement Rubric for Creativity.*

Definition of creativity:

Creativity is the capability to generate novel and appropriate outcomes by employing problem finding, resource expansion, combination searching and evaluation.

Component:

- Problem finding: discovering novel and appropriate ideas or problems that the student has never noticed before.
- Resource expansion: developing disparate appropriate resources that are novel to the individual student.
- Combination searching: synthesising selected resources by using creative combinations that are novel to the individual student and appropriate for the contextual purpose.
- Evaluation: identifying novel responses that an individual student has never proposed before, and defending the novelty of the response through sustained argument.
- Outcome: producing novel and appropriate ideas, solutions, and products.

Component	Level 0	Level 1	Level 2	Level 3	Level 4
Problem finding		Looks at scenarios from one perspective and raises superficial ideas or problems that every student can notice.	Observes scenarios from similar perspectives and recognises common ideas or problems that most students can notice.	Student(s) independently analyse scenarios from dissimilar perspectives and identify sophisticated or unusual ideas or problems ~~that are unfamiliar to peers~~.	Analyses scenarios through a broader context, thinks outside the usual boundaries and discovers novel and appropriate ideas or problems ~~that few students have noticed before~~.

Resource expansion	Recalls identical situations and lists common resources within one discipline or category, that every student knows.	Associates similar situations and collects typical resources across similar disciplines or categories, that most students know.	Uses basic skills and strategies to search and select unusual and appropriate resources across similar disciplines or categories~~, that are unfamiliar to peers~~.	Applies a variety of skills and strategies to develop disparate appropriate resources across dissimilar disciplines or categories, that are novel to the individual student.
Combination searching	Follows the same methods that students have known before in identical situations to synthesise resources.	Applies typical methods that students have known before in similar situations to synthesise resources.	Develops unusual and appropriate methods by adjusting similar approaches across similar disciplines or categories to synthesise resources~~, that are unfamiliar to peers~~.	Creates disparate appropriate methods by integrating various approaches across dissimilar disciplines or categories to synthesise resources~~, that are novel to the individual student~~.
Evaluation	Chooses a common and appropriate response that is known by every student.	Chooses a sophisticated response that most students will also identify as their solution.	Chooses an unusual and appropriate response by comparing it with peers' answers, however, the argument defending its novelty is not well sustained.	Chooses a novel and appropriate response that has not generally been proposed before, and provides arguments to support its novelty.
Outcome	Copies or restates an	Develops an idea,	Generates an idea, solution or	Creates an idea, solution or

		idea, solution or product from existing material that is mostly appropriate and contextualised.	solution or product that is a conventional combination of something existing, and/or is mostly appropriate and contextualised.	product that shows novel aspects and is appropriate and contextualised.	product that is entirely novel, highly appropriate and contextualised.

Note: Outcomes can vary, such as lab reports and essays in the written format; presentation and singing in the oral format; handicrafts, painting and robot cleaners in the format of physical products.

6.5 Chapter Summary

In the FG2 discussion, participants provided further evidence to support the validity of the measurement rubric for creativity in science. To sum up, the participants, who were experienced in teaching and/or researching in high school science, were generally satisfied with the definition of creativity and found the generic measurement rubric suitable for use in assessing creativity in science. Several amendments were supported by the participants in FG1 and FG2, including the definition of the outcome component and the performance descriptors in the rubric.

The next chapter, Chapter 7, analyses feedback from the third focus group on the updated definition, components, and measurement rubric to validate the extent to which they can be used in the physics domain. The data from two teacher development workshops are analysed in the same chapter to validate further whether teacher participants can design appropriate tasks against the defined components in the measurement rubric so that students' performance can be referenced to the performance descriptors and located on the generic measurement rubric for creativity.

Chapter 7 Validation Analysis: Physics Group Discussion and Teacher Development Workshops

7.1 Introduction

This chapter presents the results from Focus Group 3 (FG3) and two teacher development workshops. The results from FG3 provides an evidential case to answer the third research question of the extent to which a generic measurement rubric enables the measurement of creativity in the physics domain in the Chinese school setting. The result from two teacher development workshops answers the fourth research question of the extent to which physics teachers in Year 10 classrooms in China are able to develop tasks that produce evidence that can be used to locate students within a generic measurement rubric for creativity.

The defined creativity construct and associated generic measurement rubric, which have been reviewed by the first two focus groups, were critically reviewed by a third focus group (FG3) to further substantiate the evidential case for validity of the rubric for measuring creativity in Chinese high school physics students. The FG3 discussion contributed to the final version of the measurement rubric for creativity. The same participants (FG3) then took part in the design of assessment tasks for use in their classroom instruction. These tasks were designed to enable students to provide evidence that could be used to locate the students on the measurement rubric according to how much of the construct they had exhibited in their response to the task. This demonstrated how, after the measurement rubric has been constructed, it

could be used to measure creativity. In addition, the reaction of FG3 to the tasks would also provide further evidence regarding the validity, including the practicality of the use of the measurement rubric in a classroom situation.

7.2 Feedback on the Creativity Construct and Performance Standards from FG3

Creativity in this thesis is defined as "the capability to generate novel and appropriate ideas, solutions and products by employing problem finding, resource expansion, combination searching and evaluation". The participants in FG3 were six high school physics teachers in Guangdong, China, who had at least one year of high school physics teaching experience and were teaching physics in China for the years 2022 and 2023. They began by reviewing the structure and the meaning of the definition to ascertain its suitability for use in the physics domain. The distinction between creativity and other learning was then discussed by FG3 to examine whether the definition reflects the novel characteristic of creativity as perceived by the participants of FG3.

7.2.1 Reviewing the Construct of Creativity in Students

In this section, the construct regarding student creativity includes the definition of creativity and its measured components. The target of the definition and the relationship among the components are two topics involved in the discussion.

(1) The Target of the Definition

The participants struggled to understand the relationship between the

definition and the assessment. For instance, Participants P01_FG3 and P02_FG3 were unclear about the purpose of the definition and asked, “Is the definition about what teachers can do to measure students’ creativity?”, “Is the subject of the definition teachers?” Informing them that the definition describes what students know and can do in relation to the measured construct, Participant P06_FG3 suggested adding the subject in the definition as “students” to help understand the purpose of the definition and its use in practice.

In standards-referenced systems, a definition refers to a statement of the learning goals, objectives, and expectations in an area of learning (Bennett et al., 2012). It reflects intensive attention to the measurement construct other than to a specific population. The definition and rubric can be applied to anyone (e.g., teachers, lecturers, or the public) who agrees with the construct and is interested in determining how much creativity they demonstrate. In addition, it is uncommon to specify the “subject” in the definition and the performance-level descriptors in the rubric. Inventiveness in the University of Sydney’s Graduate Qualities, taken as an example, are defined as “generating novel ideas and solutions” (University of Sydney, 2022). Similarly, Sternberg (1986) defined creativity as “the ability to generate novel and appropriate ideas, solutions and products” without a specific “subject”. Therefore, there is no theoretical need to include a subject in the definition.

From the point of view of assessment theory, the definition remains the same, and there is no qualification of the subject focus. In response to this issue, the basic elements of a standards-referenced systems and their role would be introduced in the

first teacher development workshop, to supplement the assessment knowledge of FG3 participants. This idea was supported by Participant P06_FG3 as:

> The way in which generic abilities are defined and measured will be highly different from the training we receive in everyday teaching, which is aimed at students learning physics knowledge and related skills in a given year.

This comment was followed by Participant P05_FG2:

> I initially had the same confusion because the components are usually carried out by the teacher. Teachers would find and ask questions in the class, look for different combinations and teach them to students … I think what we need to do in terms of measuring creativity is going to be dramatically different from what we are used to doing, and we need additional guidance and support.

(2) Relationship among the Components

> The definition of each component is consistent with my understanding of these concepts. I think they are easy to understand … I like these definitions, especially that they include the degree of novelty in creativity and enable teachers to grasp the essential characteristics of creativity.

Taking Participant P06_FG3's above comment as a stimulus, the participants considered the components well defined and understandable. As the components can be closely related to scientific inquiry skills, two participants wondered if they could

be envisaged in a progressive relationship. Participant P02_FG3 commented:

> Scientific inquiry skills include defining problems, developing and implementing plans, drawing conclusions and evaluating the results. This is a procedural and progressive process that students will learn step by step in their lab sessions. Do these components have the same progressive relationship, i.e., do students have to be creative in finding a problem before moving on to the resource expansion component?

And Participant P06_FG3 asked:

> How can we judge the creativity of a student who is creative in problem finding, but average in resource expansion? Can he or she be creative in this case?

It's not surprising that these questions emerged in the process of exploring a creative outcome. The underlying pedagogy can be linked to the instructional technique of scaffolding in physics education, which aligns with the standards outlined in the High School Physics Curriculum, which have been largely underpinned by the levels of cognitive depth represented in the Revised Bloom's Taxonomy (Ministry of Education of People's Republic of China, 2017b). Teachers are used to unpacking the knowledge and skills embedded in the outcomes or content standards into smaller elements so that they present a progression of learning from lower to higher levels of learning, where the higher levels demonstrate more cognitively demanding use or application of the knowledge in a defined area of learning. However, these components in the measurement rubric do not represent a

sequence in the measured construct of creativity but, rather, individual (or different) but at the same time related components that together comprise the definition. They have been derived from the definition. Each component functions as a channel of observation or measurement, making possible the measurement of student creativity (as defined). A more common example is where mathematics is comprised of different components, such as, number operations, measurement, space and probability, which can be treated as individual components but they also be aggregated to produce a high-level construct called mathematics. This high-level reporting reflects what is being advanced in the case of creativity. Individual components could be reported separately as separate locations or it could be reported as creativity. This is a characteristic of analytic rubrics, discussed in Chapter 2. Students' performance in each component task can be pieced together to build a complete image of what it is each of them knows and understands about creativity. Thus, it is possible for a student to show a high level of creativity in one component but a low level in another.

After the relationship among the components and the broader construct were made clear, Participant P03_FG3 advanced that:

> Creativity consists of five components, which become a form of mesh chart for evaluation. The teacher will draw a circle for scoring in different components of the grid chart. Where students have deficiencies, targeted instruction is provided by teachers to help them progress to the next level of performance in each component.

The other participants agreed with Participant P03_FG3 that they had

provided an easy-to-understand overview of the meaning and use of a measurement rubric.

To sum up, the FG3 participants agreed with the definition of creativity and its components, and no amendment was made to the definition of the construct of creativity. The above discussion exposes teachers' unfamiliarity with the role of components when assessing a given ability, which reaffirms the need to introduce assessment knowledge related to standards referencing and its structure in the teacher development workshops.

7.2.2 Reviewing the Performance-Level Descriptors of Creativity

The participants reviewed the extent to which the performance standards can be integrated into the teaching of physics from a practical perspective using the performance-level descriptors. The participants discussed each performance-level descriptor to ensure consistent interpretation of the descriptors. By envisioning the assessment tasks for each component, the participants identified a few potential challenges in using this rubric in the classroom. With the above analysis, they answered the question of whether there is a need for a domain-specific rubric in physics.

(1) Novelty of Each Performance-Level Descriptor

The discussion started with Participant P04_FG3's comment:

> The descriptors at Level 1 are all described as "that every student can notice", which means the student who can come up with a response that

> is novel to herself or himself but known to everyone else is located at the first level. However, this student should be marked as creative by your definition.

They were unsure about the novelty of each level when interpreting the performance descriptors. In the performance standards, the statements “that every student can notice” and “that most students can notice”, which are used in Levels 1 and 2 descriptors, were intended to help distinguish between these two levels of creative performance. However, to be able to make accurate and consistent judgements about student performance, teachers should have a full understanding of each of the performance-level descriptors, rather than relying on partial descriptors for reference. In other words, the statements about what students know and can do (referring to the cognitive aspects) in relation to creativity in the first half of the descriptors should also be understood and used when referencing them to student performance. In the component of problem finding, for example, whether students’ ideas or questions manifest in their ability to analyse the situation from different perspectives is evidence to determine whether they are performing at Level 1 or Level 2. Further, the measurement rubric is not used to judge whether a student is creative or not but describes how well the student has performed on the different components of creativity. The attribute of the performance descriptors at different levels is that the higher level includes all the qualities of the lower level plus “some more” of the property.

Several examples from the participants were put forward to help them

understand and consolidate the meaning of novelty associated with different performance descriptors in measuring creativity. One of them is the story of the empty soap box that happened in a Japanese soap company, where the robotic assembly line sometimes wrapped up the paper without soap, leading to an empty packet. There were two solutions to the problem. The scientific solution was developed by the professional engineering team in that company, which spent $1,000,000 to create an X-ray scanner with a robotic arm to remove the empty box. Another simple solution was proposed by an employee in a small company, using a $25 electric fan to blow away the empty box. According to the performance descriptors in the outcome component, these two solutions can be located at the highest level because they were "entirely novel to the creators themselves and highly appropriate to the context".

The question discussed in FG3 was why the participants generally considered the second solution more creative than the first. Participant P03_FG3 said:

> Maybe one of the reasons has to do with the time and cost differences between the two solutions, with people giving preference to those simple and effective methods.

Creativity (or innovation) in business is generally tied to costs and benefits, which is different from the definition of creativity in students, where its social and economic impact is disregarded. Considering the underpinning of associative thinking in creative performance, Participant P02_FG3 proposed the following hypothesis:

> Let's put aside the costs and benefits and judge the extent of creativity in these two ideas. I think the professional engineers would be considered

> more creative if they could propose Solution B because it goes far beyond their conventional way of thinking. Similarly, the employee would be evaluated as more creative if he or she could develop the scientific method … So, what creativity requires is associative thinking, connecting different resources or ideas together using methods that are not commonly used for themselves.

Throughout the discussion of the examples, the participants developed a consistent understanding of the novelty reflected in the performance descriptors. They agreed to leave the performance-level descriptors unchanged at this stage and subsequently validate the practicality of the rubric by discussing the possibility of developing assessment tasks with different components and preconceived student responses against the performance descriptors across levels. The participants were of the view that in the process of developing the tasks and conceptualising the prospective or potential responses of students to the task, the participants believed that they may be able to gain a clearer idea of how the descriptors might be operationalised in practice and may also get ideas as to how the descriptor can be more accurately defined to capture the image of the performance.

(2) Practicality of the Performance Standards in Physics

To validate whether the performance standards of creativity can be used and embedded in a physics context, the participants drafted several assessment tasks for some of the components they were familiar with. These components were resource expansion, combination searching, and outcome. These tasks were drafted without

any assistance, based on the participants' understanding of the components and common items the participants had designed and used in their teaching. Since the participants did not fully understand the measurement rubric and its application in designing assessments, the tasks they produced were not strongly linked to creativity assessments and were not administered to the students. When writing the tasks, the participants also identified how the students might respond to the task and these potential responses from students were discussed with a focus on the extent to which the performances could be used to locate students at different performance levels on the components.

Table 7.1 shows three assessment tasks the FG3 participants drafted for the components of resource expansion, combination searching, and outcome, and the potential responses they envisaged might come from the students. While this table appears to involve a scoring process for several given tasks, it was not a scoring rubric. The tasks were not well designed to assess creativity. Also, the potential responses did not involve an in-depth discussion with reference to the performance-level descriptors so cannot be considered as exemplars or evidence of student performance at each performance level. This table shows FG3 participants' spontaneous attempt to design possible tasks that provided a stimulus for discussion that impacts on the case for validating the generic measurement rubric for creativity.

Table 7. 1 *Draft Assessment Tasks and the Potential Response from Students.*

Component	Tasks	Potential response			
		Level 1	Level 2	Level 3	Level 4
Resource expansion	How do you calculate the work done by a variable force? (P02_FG3)	Lists the methods shown in the textbook.	Collects the methods in different learning materials in physics.	Searches and selects methods in the domains of mathematics, and technology.	N/A
Combination searching	How do you measure atmospheric pressure with available instruments in a basic physics lab? (P03_FG3)	Assembles a barometer to measure atmospheric pressure in the same way as in the textbook.	N/A	Combines spring dynamometer and injection syringe to measure atmospheric pressure.	N/A
Outcome	Design a model to solve the conveyor-belt problem. (P01_FG3)	Derives an application of Newton's Second Law using the existing original problem.	Develops a sophisticated model by comparing the conveyor-belt problem to the plate problem.	Combines several physical formulas by synthesising different conveyor-belt problems.	Develops a novel model that can resolve conveyor-belt problems and plate problems and verifies the model by designing a new conveyor-belt problem.

While the outcome component was believed to be difficult to understand in

the FG2 discussion (see Section 6.2), its draft task was the fastest to be designed. Participant P01_FG3 proposed a task for the outcome component: design a model to solve the conveyor-belt problem. Even though the task needs to be improved for use in practice, and student potential responses at different levels need to be rethought if it is to contribute to the development of an associated scoring rubric, it was a pioneering attempt at thinking and measuring creativity by designing assessment tasks for each component against the rubric. However, the stimulus from the outcome task did not accelerate the task design for the components of resource expansion and combination searching. Participant P02_FG3 reflected on this situation:

> I think the components of problem finding and outcome are easy to apply in our class, but [the components of] resource expansion and combination searching are so difficult … they are related to the intermediate process that one component is closely connected to the other. In the task of measuring atmospheric pressure for the component of combination searching, students will consider useful instruments when choosing the appropriate combination of approaches. That's why we've struggled to design a tailor-made task for a particular component.

This comment reminded Participant P03_FG3 to rethink their task created for the combination-searching component, which they found inappropriate after being aware of the proper use of the measurement rubric in task development.

> … we should create a more specific task with an assigned component, not a scenario to let students free to understand and respond to it …

Participant P05_FG3 was also challenged when thinking of tasks or learning activities related to the component of combination searching with a view that:

> Creativity is a systematic capability, whereas we just concentrate on applying the rubric to the customary physics classroom mechanically, and we can't think of any good tasks for the time being. We need to be adaptable when it comes to developing and measuring creativity.

One suggestion was given by Participant P06_FG3, who connected the Internal Assessment (IA) in IB Physics to different components of creativity. One of the IA tasks in physics evaluates students' ability through writing an experimental report, which basically includes the elements of defining a problem, selecting instruments, designing an experimental device, the result and evaluation. The participants agreed that linking the rubric to scientific inquiry skills and experimental physics procedures would effectively increase the likelihood of its use in physics teaching. Participant P06_FG3 said*:*

> It is mind-challenging, all the existing assessment tasks we can think of right now definitely need to be modified to enable the measurement of creativity because we know they are not designed for creativity.

Capturing the predicted student responses to the designed tasks at different performance levels was another challenge that participants encountered in this part of the discussion. As shown in Table 7.1, several blanks indicate difficulty imagining potential responses from students to the tasks at the prescribed performance level. As advanced by Participant P02_FG3:

> This is because creativity requires interdisciplinary learning to develop associative thinking that underpins it, yet in practice, we focus only on teaching our disciplines and do not provide opportunities for students to "think outside the box".

Participant P05_FG3 commented:

> I assume that my students' performances in creativity are mainly located at Levels 1 and 2 according to the performance descriptors, and rarely reach Levels 3 and 4, and we lack a creative image in terms of student performance to inform what the creative answer to the task would be.

There are inevitable challenges in designing appropriate tasks and anticipating students' creative responses to them when measuring creativity with a measurement rubric. Nevertheless, the above practical exercise (writing a task to assess the components of the measurement rubric) and the associated discussion led to a unanimous conclusion among participants that they supported the use of a generic measurement rubric in physics contexts other than a physics-specific rubric. One reason for this is the requirement for interdisciplinary learning to be embedded in the performance standards, as two participants commented. Participant P02_FG3 said:

> The performance standards involve the integration of multiple disciplines. For example, chemical, biological, and physical knowledge is used simultaneously in response to the tasks for measuring resource expansion and combination search.

Participant P06_FG4 added:

> Scientific disciplines such as physics, chemistry and biology are experiment-based subjects and share a lot of knowledge and methods. I don't think it's necessary to specify a rubric for physics.

As a result of drafting several tasks for the discussion, Participant P04_FG3 supported using the generic measurement rubric in physics teaching:

> I also don't see the need for a specific physics rubric because we have been able to draft several tasks and assess some possible student responses in a very short period. If teachers can fully understand these descriptors in the generic rubric, they can design appropriate tasks and make professional judgements about student performance.

Participant P03_FG3, reflecting on overcoming the difficulties in the implementation of measuring creativity, suggested an alternative view by proposing a task-specific rubric rather than domain-specific.

> If we do want to create a specific rubric, one that is task-specific rather than domain-specific would be more helpful because students will perform differently according to the assessment tasks. Such a rubric could directly demonstrate what students have performed at each level, ensuring that teachers have a consistent judgement of student performance.

In conclusion, the possibility of using the generic measurement rubric for creativity in the physics domain was supported by FG3 participants, and no modifications were made to the generic measurement rubric. Informed by the

challenges mentioned above, relevant reflections and material (e.g., the meaning of measurement rubrics, the assessment tasks drafted in the discussion and from the literature) were prepared for the subsequent teacher development workshops to stimulate brainstorming on designing assessment tasks. Given Participant P03_FG3's suggestion to use a task-specific rubric, the second teacher development workshop would conclude with a discussion of the extent to which it is necessary to build a task-specific rubric for the assessment tasks to help teachers reference student responses to the performance-level descriptors.

7.2.3 Conclusion of the FG3 Validation

Being school teachers in physics, FG3 participants thoughtfully discussed the revised definition of creativity, its components and measurement rubric from the perspective of practical application in teaching. The discussion resulted in an agreement to use the construct of creativity that had been revised and validated through FG1 and FG2 discussions. The FG3 participants all agreed on the generic measurement rubric for creativity to be used in the domain-specific context of high school physics in China. Reflective barriers to understanding the measurement structure of creativity and applying the generic rubric in practice were addressed with reference materials prepared and presented at the teacher development workshops. For example, when introducing standards referencing, the role of the definitions, their components and the performance standards in the educational assessment were clearly articulated in conjunction with familiar examples of skills-based measurement in

sciences. Assessment tasks using available instruments for measuring creativity and the draft tasks shared by the participants in the previous three focus groups would be presented as part of the workshops to help understand the criteria for designing an appropriate creativity task and to inspire the creation of assessment tasks.

7.3 Feedback from the Two Teacher Development Workshops

The two teacher development workshops are the introductory workshop of the standards-referenced assessment and the implementation session of designing and determining assessment tasks in physics. The first introductory workshop is designed to provide the participants with professional support that enables them to overcome the challenges summarised from the previous focus groups. Specifically, to ensure that the participants can fully understand the meaning of standards-referencing in preparation for the task development in the second workshop, the strands and sub-strands of the Australian Science Curriculum, together with their exemplar items, were provided as an example to illustrate the meaning of the definitions, components, performance-level descriptors and tasks in assessment. The participants provided evidence of a relatively sound understanding of standards-referenced assessment by designing different tasks for the same rubric using the content from the Australian Science Curriculum (ACARA, 2018), which provided a solid foundation for promoting a cognitive shift from a focus on knowledge-driven assessment to the measurement of generic skills.

The introductory workshop also considered existing items for measuring

creativity. This was incorporated in the workshop to help participants understand the diversity of assessment tasks for each component by referencing them to the corresponding component. These tasks were used as examples to demonstrate the principles of task design for assessing higher-order thinking (HOT) (Brookhart, 2010) and used for reference to inspire participants' thinking when designing appropriate assessment tasks. The workshops provided further evidence to support the validity of the measurement rubric by teachers being able to design appropriate assessment tasks for use in their classes after peer-reviewed discussions. In envisioning possible student responses to the tasks they designed, participants assessed the feasibility of establishing task-specific rubrics as proposed by FG3.

7.3.1 Designing Assessment Tasks for Each Component of Creativity

After the introductory workshop, the participants were asked to develop appropriate assessment tasks for each of the components in the generic measurement rubric in the second teacher development workshop. These tasks were drafted by first referring to the item format of existing creativity assessment tests (i.e., open-ended questions used in Torrance's Test of Creative Thinking, Evaluation of Potential Creativity and Test for Creative Thinking-Drawing Production) and adapting them to physics teaching by adding physics scenarios and content. During the peer-review sections (the Appendix G for the agenda of the workshops), the participants critically examined the extent to which the draft task was able to assess the intended component in the measurement rubric and the extent to which the students' responses to the task

were able to be referenced to the performance-level descriptors. This second workshop provided sufficient time for the participating teachers to revise and redesign their draft tasks based on the peer-review feedback, so they could finalise the tasks they believed were appropriate to assess creativity in their practice and were able to locate different students' responses to the tasks on the performance standards. The assessment tasks designed by the participating teachers, as well as the peer-review discussion, provide evidence of the validity of using the generic measurement rubric for creativity in the specific domain of high school physics in China.

The final version of the assessment tasks for the components of problem finding, resource expansion, combination searching, evaluation and outcome created by the participants is shown in Table 7.2, indicating that the generic creativity rubric can be applied in the physics context. The participants in the workshops were represented by the code starting with the letter P as a participant and two digits for the participant number, followed by the letter W referring to the Workshop.

Table 7. 2 *Final Versions of Participants' Assessment Tasks.*

Component	Assessment task
Problem finding (plus evaluation)	P01_W: Given a photo of people travelling with masks, please ask as many physics questions as possible that are novel and appropriate.
	P02_W: Please list as many unique disadvantages of laboratory ammeters as possible. (Evaluation: Which disadvantage do you think is the most creative to you and why?)
	P03_W: Given a picture of two horses pulling an iron ball in the middle, please ask as many physics questions as possible that are

	novel and appropriate.
	P04_W: Please observe the working process of a “Jumping frog” machine; what are the shortcomings of this machine?
Resource expansion (plus evaluation)	P01_W: Please write down novel instruments or electrical components to measure body temperature by searching and screening. (Evaluation: Which instrument do you think is the most creative? What considerations have you made in your choice?)
	P03_W: What novel and appropriate materials are you going to use to measure atmospheric pressure? (Evaluation: Which material do you think is the most creative for you? How did you decide on these materials?)
	P05_W: Please find as many novel ways as possible to measure the local acceleration of gravity.
Combination searching (plus evaluation)	P05_W: “Paper bridge” – please design several novel paper bridges using only A4 paper. You can fold or shape the paper, and the bridges should be able to hold the weight of at least one egg. (Evaluation: Which design do you think is the most creative to you and why?)
	P06_W: “Cardboard tower” – please connect the playing cards in several unique ways to build a tall tower. (Evaluation: Which way of combination do you think is the most creative to you and why?)
Outcome	P01_W: Please design a unique and light mask to protect against viruses.
	P02_W: Please design several novel plans to measure atmospheric pressure.
	P04_W: Please write your proposal for improving the “Jumping frog” machine in Task 1, which should be novel and able to overcome the shortcomings.
	P05_W: “Egg drop” – please design and build several novel constructs to keep a raw egg from cracking when dropped from ever-increasing elevations.

	P06_W: Please design and build your own novel optical instruments with a convex lens.

An unavoidable challenge faced by participants in designing assessment tasks is to differentiate between the knowledge-oriented tasks that participants are used to in their daily physics teaching and the capability-oriented (in this case, creativity-oriented) tasks that are currently required. In the former task, students' responses are centred on the scope of knowledge closely related to the syllabus with "one right answer", while the latter task required students to be open-minded and give a variety of answers. This difference was magnified when designing tasks for the component of combination searching. For example, participants considered the draft task, "Please design several circuits to measure resistance using the electrical components provided by the teacher", inappropriate in measuring creativity because the electrical components provided by the teacher generally came from school laboratories and were used to teach the conventional ways of combining circuits in textbooks (series and parallel circuits). Therefore, students who were given this task could only recall and construct series or parallel circuits taught in class as different responses to the given questions.

FG3's advice on designing interdisciplinary tasks inspired participants to link existing STEM-based school activities to the assessment tasks in the combination-searching component. As can be seen in Table 7.3, "Paper bridge" and "Playing card tower", proposed by Participants P05_W and P06_W, were two STEM activities that were conducted in two senior high schools in China, where students were asked to search different ways of combination to fold and combine the given materials (e.g.,

paper or card) for a particular purpose. In contrast to tangible outcomes, such as a final physical paper bridge or cardboard tower, these two tasks in the component of combination searching assess the intermediate process before students come up with a product. These tasks can be flexibly measured in an oral or drawn format appropriate to the instructional setting.

The assessment tasks described above reflect the nature of minimising or even breaking down disciplinary boundaries in measuring student creativity, which is consistent with the defined generic construct of creativity. These evidential tasks reinforce the idea that a generic measurement rubric is applicable and appropriate in the teaching and learning of physics if teachers are willing to undergo considerable training to understand the generic measurement rubric and be able to design assessment tasks that are well aligned with the rubric.

7.3.2 Reconsidering the Need for Task-Specific Rubrics for Creativity

The participants in the second workshop reconsidered their suggestions from their focus group (FG3) discussion on the need for task-specific rubrics in the light of the final tasks they had designed. They discussed the extent to which a task-specific rubric could contribute to a consistent interpretation and accurate reference of student responses to performance-level descriptors in the rubric and improve the reliability of the assessment. Recognition of the strengths of a task-specific rubric and its widespread use in standardised testing across disciplines did not support its establishment among the participants. This discussion segment evolves into two main

reasons to reject the proposed task-specific rubric: huge workload and lack of student response samples.

> I think the generic rubric is understandable, it may be great if we could have a task-specific rubric to specify and contextualise the meaning of the performance descriptors. However, it would be overload for the teacher to be responsible for designing the task and creating task-specific rubrics.

This comment was supported by P02_W, who compared task-specific rubrics in measuring physical knowledge and creativity, and arriving at favouring the use of a generic measurement rubric in physics.

> It would be a huge workload for us teachers to create different task-specific rubrics. Knowledge-based tasks are well structured to limit student responses to specific knowledge, and we can easily predict what students will respond to when designing the task. Creative tasks, however, are quite different; they allow for divergent thinking and welcome unique and novel responses that even teachers cannot anticipate when designing a task. Maintaining generic descriptors in the rubric is more useful and adaptable in situations fraught with uncertainty.

Other participants suggested implementing several pilot tests for these tasks to fill in the gaps in the student response sample and assist in establishing task-specific rubrics in the future. Participant P06_W noted that the lack of adequate response samples may lead to inconsistencies between teachers' expectations and students'

actual responses, making task-specific rubrics impractical and meaningless. If task-specific rubrics were deemed necessary, Participant P05_W suggested pilot testing to obtain a larger sample of student responses, which would enrich teachers' images of students' actual performance in terms of creativity. Participant P04_W endorsed the feasibility of this proposal by conceptualising the process of building standard examples in task-specific rubrics, stating:

> For people like us who know the rubric and descriptors well, we don't need task-specific rubrics … I agree that the standard sample of student responses can be collected through several pilot tests. Some typical answers can be attached as samples to the corresponding descriptors as samples … It's like what we do when we grade papers. We stamp different scores on some typical responses and synthesise these responses into a standard sample to help speed up the scoring process and improve inter-rater reliability. We can design standardised examples and task-specific rubrics when we have more pilot tests and enough student responses.

However, growth in student creativity can only be measured and monitored if the assessment tasks and their task-specific rubrics were carefully designed to align well with the generic measurement rubric. This brings us back to the requirements for the participants to fully understand the construct of creativity defined in this thesis, and the associated performance standards that constitute the generic measurement rubric, before designing assessment activities to measure creativity. Therefore, the

participants unanimously agreed that by providing tailored professional support, teachers would be able to understand the generic measurement rubric for creativity and design appropriate tasks that are aligned with the rubric with maximum flexibility to measure students' creativity in physics contexts.

7.3.3 Conclusion of the Two Teacher Development Workshops

Participants' success in designing assessment tasks for different components of creativity validates the applicability of a generic measurement rubric for creativity used in the domain-specific context of physics. The utilisation of STEM school programs in creativity assessment indicates the interdisciplinary learning inherent in creativity, which reflects the feasibility of using the generic measurement rubric in specific domains without the need to define a domain-specific rubric. The endorsement of using a generic measurement rubric in physics was reiterated in the discussion of establishing task-specific rubrics. The desire for novel and unexpected student responses leads to the occurrence of unpopular workloads and an incomplete impression of students' progressive performance on the task when establishing the aligned task-specific rubrics. The physics teachers who participated in the workshops prioritised a thorough understanding of the performance-level descriptors over building task-specific rubrics so that assessment tasks could be easily designed, and student responses could be accurately referenced to the performance-level descriptors. In other words, they saw no need to create a task-specific rubric that is aligned with a generic measurement rubric when measuring student creativity in physics.

7.4 Chapter Summary

This chapter presents the validity evidence from the FG3 discussion and two teacher development workshops regarding the appropriateness of a generic measurement rubric of creativity as applied to physics. The participants' feedback on reviewing the generic measurement rubric in the third focus group discussion, and the assessment tasks created with professional support in the two workshops, provided an evidential case to support the effective use of a generic measurement rubric for creativity in the physics context in a Chinese high school setting.

In the next chapter, the validation results from the three focus groups and two teacher development workshops presented in Chapters 5 to 7 are summarised to address the research questions of this thesis. Chapter 8 also discusses the implications and limitations based on the methodology used in this thesis.

Chapter 8 Discussion and Conclusion

8.1 Introduction

This chapter discusses the results relating to the four research questions regarding constructing and validating a measurement rubric for assessing student creativity. The implications and limitations of this thesis and the resulting suggestions for further research using modern measurement theory and standards-referenced assessments as a basis for producing measurement rubrics for creativity are also presented. The chapter concludes by describing the possibility of generalising the process of developing a measurement rubric through this research. This thesis shows how the process of generating a measure of growth in creativity can be generalised to other 21st-century skills or generic capabilities.

8.2 Research Question 1: What is an appropriate measurement rubric to measure incremental growth along the defined components of creativity?

8.2.1 Discussion on the Defined Components of Student Creativity

The underlying theory for describing and defining student creativity in this thesis is Kaufman's Four-C Model, which provides a developmental trajectory of creativity over an individual's lifetime. The "mini-c" in the model regards students' progress in learning as a form of creativity and defines the extent of novelty at an individual level, regardless of social influence. Concentration on the novelty that is original to the creators (students) themselves implies that the assessment of creativity

in students should be developmentally oriented and that the defined components should be teachable and observable in everyday learning (Claxton & Lucas, 2013; Lucas, 2016; Sternberg, 1996). With this consideration, the definition of creativity and its components are concentrated on the cognitive factors in the generation process of a creative outcome.

The overall structure of the definition incorporates both creative potential and outcome, and the selected components received full support from the participants from the three focus group (FG) discussions. The FGs reviewed the definition, components and performance-level descriptors representing the performance standards describing growth in creativity as defined for the purpose of this thesis. The intention of the FG sessions was to provide evidence to support the case for the validation of the measurement rubric. They appraised the structure of the definition as "being consistent with the journey of creativity", "reflecting the fact in a teaching-and-learning setting", "indicating both ends of the scale from what is capable and what should happen as the outcome" and "observable in the classroom". These comments are consistent with the emphasis in the literature (Brookhart, 2010; Kaufman & Beghetto, 2009; Starko, 2013; Vincent-Lancrin et al., 2019), which gives confidence that the measurement rubric is measuring the construct it is supposed to measure. Although the creative process, represented by creative thinking, is of general interest in creative education in schools (Lucas et al., 2023), the measurement of student creativity should include both the creative process and the outcome (Barbot et al., 2011; Barbot et al., 2015; Lubart et al., 2013). The reason for this is that focusing

only on creative potential without counting its final achievement is incomplete in the development of creativity. On the other hand, focusing only on creative outcome is sometimes too difficult for school-aged students to achieve (McClure & Jaeger, 2020; Starko, 2013; Sternberg, 2015). Similarly, participants' comments on the defined components (i.e., problem finding, resource expansion, combination searching, evaluation and outcome), such as "high consistency between the definition and components" and "workable in the classroom", indicated that they believed that the components selected were essential and aligned with the construct of creativity, and were potentially assessable in a classroom situation. A high degree of alignment between the definition of creativity and its components is fundamental to subsequent effective assessments as they explicitly define what is being measured and further contribute to the use of assessment results and their legitimate interpretation (Cizek, 2020).

While the definition of creativity, the related components and the performance standards developed in this thesis are generic, to be used across different domains and schools, it is not the intention of this thesis to develop a definition and components of creativity that are accepted globally. Rather, the methodology developed and presented in this thesis is intended to provide a concrete example of how a generic measurement rubric for creativity can be defined to help teachers and schools identify and assess student creativity in their current school-setting.

A very important difference between this thesis and the way a more "omnibus" measure of creativity would be developed is that the latter would generally have a

requirement of obtaining validation evidence from “experts” across the broader population that would be using the measurement rubric. In addition, the general intention is that when such measurement rubrics are developed for use across a broader population, the intention is to compare results across the total population, and so the assessment of students would generally be standardised. This generally means that the assessment method involves a standardised test (e.g., Programme for International Students Assessment 2018 for global competence, OECD, 2018a, 2018b). A feature of the measurement rubric developed through this research is that the assessment is left to the teachers themselves. The resulting student output (evidence) from the assessment is used to locate the students along the measurement rubric. This means that teachers can use different assessments to locate the students on the measurement rubric and still compare their performance against the same performance standards. The purpose of the measurement in this thesis is very much school- and teacher-focused rather than accountability-focused.

Given the multidimensionality of creativity, other influential factors, such as personality and motivation for creativity, could be incorporated into the definition components. For example, the current definition of creativity may be inappropriate and would need to be redefined if the educational context advocates the cultivation of the five creative habits (imaginative, inquisitive, persistent, collaborative and disciplined) in primary and secondary schools (Lucas, 2016; Lucas et al., 2013).

8.2.2 Discussion on the Performance Standards of Creativity

The performance standards of creativity developed in this thesis are a sequence of performance-level descriptors designed to express differences in levels of performance on the components that evolve from the definition of creativity in terms of what it is students know and can do. Performance standards have been commonly used in measuring knowledge, skills and capabilities because of the recognised advantages documented in the literature, including informing student learning, monitoring student progress, and supporting teacher judgements (Black & Wiliam, 2018; Briggs & Furtak, 2019; Brookhart, 2018; Brookhart & McMillan, 2020; Cizek, 2020; Shepard et al., 2018; Wilson & Lehrer, 2021).

However, they are rarely used when assessing creativity. A hindering factor to its application relates to the paradox between creativity's "uncertainty" and the meaning of "standard" (Beghetto, 2019; McClure & Jaeger, 2020). While creativity is considered unpredictable, "standards" are established based on a set of predetermined outcomes. This contradiction can be addressed, hopefully resolved, if the performance standards are not for standardising and describing what a specific creative response looks like, but rather focus on the characteristics of creative performance (i.e., novelty and appropriateness) and what students should know and can do to provide evidence that can be aligned with one of the performance levels for each of the components assessed. The measurement rubric for creativity drafted in this thesis, which describes growth in each of the component of creativity through performance-level descriptors is indirectly validated as addressing this contradiction. According to the feedback

from the three focus groups, the revised performance-level descriptors clearly demonstrate the cognitive differences between levels, thus showing incremental growth in the components of creativity. Such growth enabled the participants to relate and reference a variety of student performances to performance-level descriptors when reviewing the practicality of the generic measurement rubric in their teaching contexts.

This feedback provides further evidence that the components are aligned with the construct (creativity), and this adds content and construct validity evidence to the evidential argument supporting the validation of the measurement rubric. The performance standards define what students know and can do at each level of performance in various components. Student performance is aligned to these standards through the teacher (student or peer producing the evidence when completing the task and the result being scored against the rubric) providing more information for teachers, students and others that can be used to better develop students' creativity compared with what would normally be provided through traditional normative scoring. This measurement rubric provides an image of growth in creativity so that student performance can be located on the developmental continuum of creativity, thereby emphasising what is needed to enhance creativity and, as such, provides direct information for teachers and students to learn.

Some of the benefits of having performance referenced to standards nominated by the three FG discussions include "teachers making professional judgements about students' creative performance by referencing student responses

and submissions to the performance standards". This then "enables teachers to develop learning strategies for assisting students in moving along the developmental continuum and becoming more creative". Ultimately, teachers can then "assess the effectiveness of creativity education by monitoring students' learning progress in relation to creativity".

The FG2 participants did express some concerns about the tension between using normative information and terminology inherent in creativity and the description of the "image" (of what it is a student at this level knows and can do) that characterises the performance-level descriptors that comprise the performance standards. The judgement of novelty tends to traditionally imply a comparison to a normative group of students; that is, it is novel compared with what would be expected from other students. On the contrary, standards referencing refers to the process of giving meaning to student achievement by referencing it to a set of predetermined standards of performance (Tognolini & Davidson, 2012). This is generally understood as an absolute judgement made against pre-determined standards without reference to the performance of others. However, the "ground-up" construction of standards relies on practical considerations of retrieving what is known to happen and what can be achieved in each domain (Cizek, 2012; Sadler, 1987). In practice, determining standards for any construct requires teachers to initially describe how well students are doing in terms of cognitive development and what they are expected to achieve relative to a normative student population (Hammer, 2007; Hendry et al., 2012; Suskie, 2018). For example, a group of

experienced teachers or educational specialists will be asked to imagine what an advanced student (e.g., in the top 5 per cent of a group of students) knows and can do. Statements and characteristics of different levels of students (image of the "norm") will be summarised as standards to show incremental growth in learning. In a philosophical sense, standards have a normative character that requires continuous validation and updating through the collection of new student data and the involvement of larger educational segments. It would be fair to say that there is a standard under a norm. However, the grading of the students is not limited to the ranked percentage on which the norm is based. With this consideration, it would be more appropriate to describe standards referencing as a stable reference rather than an objective or absolute measurement (Sadler, 1987).

The reflective words and phrases in the initial rubric that were commented on as implying normative comparison in the performance standards, such as "common" and "most students know" and their synonyms, triggered a lively discussion in the three FGs. The resulting solution was to add more contextual descriptions to demonstrate the cognitive depth of what students are thought to have achieved in that "common" or "uncommon" performance while maintaining the use of those "contentious", comparative but necessary words. On the one hand, similar "normative" words are an important factor in distinguishing levels of performance in terms of creativity. For example, they have been used in creativity-related measurements such as the OECD's assessment rubric for creativity (Vincent-Lancrin et al., 2019) and the PISA creative thinking framework (OECD, 2019).

The performance standards in this thesis were structured to be inherently aligned with the definition of creativity for students, arguing that the definition provides an image of growth in developing creativity that is appropriate and achievable through everyday teaching and learning and not developed necessarily to socialise and universalise a developmental theory of creativity on a global scale. The process of developing performance standards demonstrated in this research was considered by the FGs as providing a practical example of applying standards-referenced assessment in creativity. It can be transferred to different educational settings to develop school-based measurement rubrics that are appropriate for the contextual purposes that exist within the school. To that extent it provides an appropriate rubric to measure and show incremental growth in the components of student creativity as defined for the purpose of this thesis.

The next section considers the evidence provided by the FGs to support the validity of using the generic measurement rubric to measure growth of creativity across disciplines.

8.3 Research Question 2: To what extent does a generic measurement rubric for creativity enable the measurement of growth in creativity across disciplines?

As one of the "21st-century" skills, creativity has been considered a generic capability that can be cultivated through different disciplines and contexts (Beghetto & Sriraman, 2017; Patston et al., 2021; Vincent-Lancrin et al., 2019). With this understanding, the measurement rubric in this thesis is developed to be generic and, as

such, has been designed to be generalisable across different disciplinary contexts. Some studies advocate constructing and using domain-specific instruments and rubrics by contextualising the generic measurement rubric to better fit the disciplinary context and ensure more accurate judgements in the assessment (Kaufman et al., 2008; Lubart et al., 2019; Lubart, 2019). The validation process took this consideration into account by inviting participants from a variety of disciplinary backgrounds (i.e., linguistics, history, science, technology, engineering, arts and mathematics) in FG1 and science teachers in FG2 to share comments on using this generic measurement rubric in their disciplinary contexts. The outcomes from the FG1 and FG2 deliberations reached an agreement on the feasibility of using a generic measurement rubric for measuring creativity. They eventually converged on the idea of "thinking outside the box" (i.e., jumping out of disciplinary boundaries in teaching and learning), which they considered to be a characteristic of creativity and could best be met by using a generic measurement rubric. They valued interdisciplinary or transdisciplinary learning as critical to teaching and assessing creativity in schooling. This outcome is generally supported as consistent in the literature, that interdisciplinary or transdisciplinary learning, particularly in the context of STEAM education (science, technology, engineering, arts and mathematics), is effective in enhancing creativity and thinking skills (Burnard et al., 2022; Burnard et al., 2021; Perignat & Katz-Buonincontro, 2019; Wilson et al., 2021).

In addition, using a generic measurement rubric enables opportunities for designing tasks and assessing them in various ways across disciplinary boundaries.

Based on participants' feedback in FG 1 and 2, the measurement rubric developed in this thesis allows for individual and collaborative assessments (e.g., an individual task of building a card bridge and a collaborative project of designing a water rocket). Using a generic measurement rubric enables its integration with other rubrics to measure a school-level project. For example, a school project to "Create Your Vacuuming Robot" could be designed to measure creativity and other STEAM-related components that may extend beyond the original definition of creativity. These components may include the outcome component from the measurement rubric for creativity, engineering functionality, scientific vacuuming ability and artistic aesthetics.

The process of validating the measurement rubric for creativity revolved around two main discussion foci: the extent to which creativity was clearly defined and the extent to which the performance standards clearly differentiated between performance levels calibrating each of the components. According to FGs 1 and 2, an ideal definition of creativity for a teacher should describe what students know and be able to do in terms of creativity, using simple, straightforward, positive and easily understandable language. The participants' requirement for a clear definition of creativity is consistent with what is documented in the literature (Brookhart, 2010, 2013b; Lucas, 2016). This is also consistent with measurement theory in the physical science, where the process of building the measuring instrument is carried out first and then, once the instrument has been developed, it is used to measure the relevant attribute (e.g., height, weight, heat). Providing a direct image of what it is intended to

measure as expressed in the measurement rubric (creativity in this case) elaborates the direction and purpose of the subsequent assessment design. In accordance with participants' feedback, the definition of creativity was amended accordingly, to avoid any broad and abstract concepts, and the definition of each component was revised to clarify its expected performance at the student level. These modifications were discussed in FG2 and FG3, with participants commenting that the updated definitions of creativity and its defined components were better than the original and had a higher degree of constructive consistency. That is, this instance of the measurement rubric should be more valid than the previous instance. Directly describing what students should be doing in relation to creativity and its components can reinforce the link between the definition and teachers' disciplinary background, and facilitates their thinking about possible assessment tasks.

The second focus, on the distinction between performance standards, concerns the judgement of accuracy and consistency in assessment. As introduced in Chapter 3, the performance standards are the integration of the Revised Bloom's Taxonomy, and Structure of Observed Learning Outcome (SOLO) Taxonomy in terms of the quality reflected in the task response, and the Associative Theory of Creativity to indicate the discrimination of performance across performance standard levels. These taxonomies and theories were intertwined to holistically demonstrate the cognitive depth in performance standards. One suggestion from the FG2 discussion reflects participants' advanced understanding of the Revised Bloom's Taxonomy and their proficiency in applying the rubric in practice. FG2 participants suggested using different verbs in the

Revised Bloom's Taxonomy to represent and visualise the cognitive growth in performance standards so that a clear growth path along the developmental continuum in creativity can be described by the performance standards. This makes a degree of intuitive sense as the Revised Bloom's Taxonomy has been widely used in educational curricula and syllabi development to describe the development of knowledge and skills in each disciplinary domain (Chandio et al., 2021), such as in the *Scheme on Senior High School's Curriculum* (Ministry of Education of People's Republic of China, 2017a) and the NSW Education Standards Authority syllabuses (NESA, 2021). Teachers are familiar with planning and implementing effective teaching and learning based on the learning outcomes in the syllabus. They can evaluate the extent to which students have achieved the learning outcomes by designing and using appropriate test items (Chandio et al., 2021). With this reality in mind, the verbs characterising the different levels of the Revised Bloom's Taxonomy were used in the measurement rubric to make it easier for teachers to understand the growth in cognitive processing, defining the progression along each component.

The discussions in FG1 and FG2 noted in this thesis demonstrate a detailed process of how the draft measurement rubric for creativity can be validated and improved, based on participants' comments. At the end of the FG processes, the evidence collected provides support for the substantive argument that the rubric is measuring what it is supposed to measure. It is noticeable that the sample sizes of FG1 and FG2 were relatively small, and the resulting agreement of using a generic measurement rubric for creativity across disciplines is not generalisable beyond these

two focus groups. If the intention is to use the measurement rubric more broadly than the disciplinary and cultural representations in those two FGs, then a bigger and more representative sample of participants would need to be involved in the validation process. In this research, a third representative group was established to represent the views of the teachers in schools in China who would use the measurement rubrics. It should be stressed that in this thesis, the participants of FG3 had little experience with or knowledge of creativity. This limits the value of the evaluative commentary but does sign-post a process for building a validation case going forward.

In some cases, domain-specific rubrics may still be needed to help teachers make accurate judgements when measuring creativity in specific disciplines. If this is the case, a domain-specific rubric can be developed following the process demonstrated in this thesis. A domain-specific rubric should be aligned with the generic rubric, which means the evidence of student performance from a domain-specific rubric can be referenced to the generic rubric. A practical example where the issue of domain-specific or generic measurement rubrics might apply is in science. Science could comprise different discipline areas (courses) within the subject of science (e.g., physics, chemistry, biology). The question would be: Do teachers need to have domain-specific performance standards (i.e., the standards impregnated with words that reflect the various disciplines) or can teachers use the more generic descriptions associated with the performance standards to make a judgement of the location on the measurement rubric? The fact that the participants in FG2 have indicated that they can work with the generic measurement rubric suggests that

although the teachers in the different courses teach different content, their results can be directly compared because the performance is “mapped” onto the same measurement rubric. This is potentially a very powerful finding from this thesis and should be the focus of future research.

The next section addresses the third research question regarding the extent to which Chinese teachers are of the view that they can effectively use a generic measurement rubric for creativity to measure the performance of students in creativity in physics in a Chinese high school.

8.4 Research Question 3: To what extent does a generic measurement rubric enable the measurement of creativity in the physics domain in the Chinese school setting?

The participants in FG1 and FG2 comprised high school teachers in Sydney, Australia, a context that differs from the Chinese school setting that provides the context for this thesis (physics). Given these two countries’ different cultural and educational contexts, the measurement rubric needed further validity evidence to be collected and interpreted. In this case, the third focus group (FG3) involved selecting and working with six Chinese physics teachers. As seen in Chapter 7, there was no further amendment to the measurement rubric for creativity after FG2, and FG3 participants unanimously agreed to use the generic rubric to measure creativity in the physics domain. They argued that physics, as a branch of the science domain, shares the same scientific methods, inquiry skills, and experimental principles as other

scientific subjects such as biology and chemistry. Additional validity evidence supporting the credibility of the measurement rubric is provided by the assessment tasks designed by FG3 participants. These tasks were a first draft and needed further discussion in the following teacher development workshops. They provided preliminary evidence to support the view that each component of the defined creativity is observable in a physics classroom as the FG3 teachers constructed assessment tasks for the components of resource-expansion, combination-searching and outcomes of the creativity measurement rubric.

Having teachers as participants in this thesis has dual benefits: collecting their feedback as validity evidence of the creativity measurement rubric and informing the development of tailored professional activities to enhance their professional development in accord with the evidence that they provided during the FG activity. FG3 participants made several explicit challenges in using a generic measurement rubric for measuring creativity in their classroom. These challenges include (1) understanding the nature and structure of the measurement rubric and the construct of creativity, (2) designing assessment tasks that provided appropriate information (evidence) that could be used to locate students on the measurement rubric and (3) referencing student performance to the performance standards that differentiate levels of performance on the rubric.

Understanding the measurement rubric requires teachers to invest more effort in understanding the definition of creativity, the components, and their meanings, and most importantly, the meaning associated with performance-level descriptors in the

performance standards that show the growth of creativity along the developmental continuum. The apparent lack of understanding of the measurement rubric shown through the FG3 discussion can be contributed to two factors. One is that the measurement rubric is generic and not specific to any particular discipline or even assessment task. Teachers must autonomously embed the definitions and performance standards in their teaching context. The other factor is that creativity is not prioritised in classroom teaching and assessment in China due to the lack of explicit curricula associated with creativity in school syllabi. This leads teachers to rely on their "gut feeling" to judge student performance in creativity when there is a creativity-related task or challenge. A generic measurement rubric for creativity with a strong theoretical base provides teachers with a systematic and detailed image of what creativity means for students and how much is expected to be achieved at each performance standard level. Once teachers from different disciplinary backgrounds agree to use the same measurement rubric for creativity, they can collaboratively discuss and assess creativity with the shared language and standards rather than the "gut feeling" that varies from one case to the next.

Designing assessment tasks against the measurement rubric is the second challenge for teachers, even though the teachers from FG3 indicated that the measurement rubric was flexible and adaptable to different learning activities (e.g., the routine laboratory class and STEM activities that have been organised in their schools). In traditional classroom assessments, teachers deliver prompt scores or informative feedback on a specific assignment or test using a scoring (task-specific)

rubric (Hafner & Hafner, 2003; Jönsson & Svingby, 2007; Mertler, 2001). A scoring rubric is a set of scoring criteria delineated according to the specific scenario of the task or item being tested (Mertler, 2001; Moskal, 2000). In contrast, the measurement rubric developed for this thesis was predetermined using information from the literature review and a series of validation processes, demonstrating the measurement construct of creativity in general. In this case, teachers were required to design tasks that were appropriate to their teaching context and aligned with the measurement rubric for creativity. The ideological challenge for teachers is the misalignment between the measurement focus on creativity in the rubric and the dominant disciplinary examinations that focus on knowledge and skills in a subject area (Beghetto & Glăveanu, 2022). Therefore, teachers are generally unable to draw on their experience in subject assessments to design tasks for measuring creativity against the measurement rubric. The misalignment makes explicit the measurement challenge regarding the measurement of "intermediate process" components (i.e., resource expansion, combination searching).

Referencing student responses to the performance standards is considered the third challenge for teachers to apply the generic measurement rubric in classroom assessments. As discussed above, teachers are generally more concerned with students' progress in subject knowledge and skills, and they tend to lack the skill associated with investigating and collecting creative performances and behaviours in the classroom. As a consequence of teacher challenges in writing items aligned with the components and performance-levels descriptors in FG3, a process was developed

to raise teacher awareness of how such items could be written. After attempting to draft several assessment tasks to provide evidence of student performance against the generic measurement rubric in the FG3 discussion, the participants agreed that they found the task difficult. They suggested that they lacked sufficient developed "images" (impressions) of students' creative performance that could be aligned with the performance-level standards and that had difficulty conceptualising how students would respond to various prompts when drafting an assessment task. The participants argued that it would be challenging for teachers to reference students' responses to the performance standards and locate them on the generic measurement rubric with any degree of confidence. This is not a surprise because it is a process that Chinese teachers have generally never experienced. Traditionally, teachers set tasks and know the answers to the task. The students know they get rewarded from generating the answer the teacher has already decided upon. This common practice, of course, inhibits higher order thinking in students.

The next section specifically discusses the process by which the participants designed the assessment tasks and received tailored professional support during the teacher development workshops to address the reflective challenges mentioned above.

8.5 Research Question 4: To what extent can physics teachers in Year 10 classrooms in China develop tasks that produce evidence that can be used to locate students within a generic measurement rubric for creativity?

8.5.1 Professional Support Provided in the Teacher Development Workshops

The importance of providing professional support to teachers and educators to help them shift from traditional knowledge- and skills-based education to an education that helps prioritise the development of 21st-century skills through the provision of clear learning outcomes, rubrics, pedagogies and assessment materials has been recognised (Lucas, 2019; OECD, 2019). In this thesis, two teacher development workshops took place to support teachers in designing tasks that provided appropriate evidence for locating students on the measurement rubric. These two workshops included three participants from FG3 and three additional participants who met the same eligibility criteria as for FG3. Previous research had pointed to insufficient knowledge of creativity among physics teachers (Xu, 2018) as a reason for not being able to develop tasks to measure it. Therefore, the first workshop was initially designed as an introductory session to provide conceptual knowledge about creativity by introducing several significant creativity theories, such as the developmental Four-C Model (Kaufman & Beghetto, 2009, 2013) and the Associative Theory of Creativity (Mednick, 1962). After feedback from the participants involved in the FG3 discussion, the workshop was redesigned to better address a challenge associated with the teachers' lack of understanding of measurement rubrics (see Appendix G for the agendas of the two workshops). Given that these rubrics are based on modern measurement theory and standards-referenced assessment, abbreviated units associated with modern measurement theory were introduced into the content of the first workshop to make clear the difference between scoring using a traditional marking guide or task-specific rubric and measurement rubrics with characteristics

like those developed through this research. The focus was very much on showing how these measurement rubrics measure the growth of a construct, such as creativity, using performance-level descriptors to define growth or progress (Bennett, 2015; Bennett, 2018; Black et al., 2011; Masters & Forster, 1996; Shepard et al., 2018; Wilson, 2018; Wilson & Sloane, 2000). A unit that also addresses the characteristics of a standards-referenced model, which makes explicit the link between standards and the measurement model, was also introduced to the workshop to make clear that designing good-quality assessment tasks requires a direct link between the task, the performance the task elicits (evidence) from the students and the performance level for each of the components in the measurement rubric. It also makes clear how the performance of the students to the tasks is used as evidence to locate the student in the rubric using a pairwise comparative process. At the end of the first workshop, participants were able to provide sound explanations and exemplar tasks as per the principles for assessing higher order thinking (Brookhart, 2010) to demonstrate the sophistication of their task design in the context of physics.

A second workshop was organised for the same teachers, who designed assessments using the knowledge promulgated in the first workshop. While they understood the difference between scoring and a measurement rubric, their preference for assessing knowledge and skills in a specific subject area persisted. This hindered their ability to design tasks that assess intermediate process components (i.e., resource expansion, combination searching). Teachers were given tailored support to shift their thinking and break down disciplinary mindsets by discussing creativity assessments in

various fields, including art, writing, and STEAM programs. These assessments were drawn from the literature, practical examples shared in previous focus groups and content posted in school newsletters. During the second workshop, the teachers' tasks were individually reviewed and given feedback by the researcher and group peers based on predetermined assessment criteria. At the end of the second workshop, teachers were able to design tasks that were more aligned to the measurement rubric, and to discuss and consistently reference potential responses (as evidence) of students to the performance standards of creativity. These achievements demonstrate their increasing understanding of the measurement rubric for creativity. It suggested that the tasks they had written using the knowledge and understanding obtained from the first workshop and personalised feedback from the second workshop were better aligned with each of the components that comprise the measurement rubric. This means that the result obtained from such tasks are more valid in relation to measuring the construct of creativity as defined for this study.

Student responses to the designed tasks were not collected in this research, leaving the challenge of referencing student performance standards unaddressed. This could be the next stage to extend the results of this thesis to operational implementation in real classroom learning. According to education assessment, a task-specific rubric (or scoring rubric) could be helpful in this circumstance as it contextualises the meaning of each performance descriptor to each specific task for reliable scoring purposes (Hafner & Hafner, 2003; Mertler, 2001). However, teachers in the second workshop rejected this proposal due to the huge amount of work that

this would entail if teachers were to develop such task-specific rubrics. Moreover, they were temporarily unable to design a task-specific rubric because of the insufficient evidence they had of what students' creative performance might entail through their daily teaching.

Teachers from the second workshop called for ongoing professional support in the implementation process: referencing students' responses to the measurement rubric, establishing task-specific rubrics (if necessary) and analysing and reporting data. Once the student responses are collected, teachers will encounter the third challenge of referencing these responses to the performance standards. Organising a workshop or group discussion can create a dynamic environment where multiple interpretations of student responses can be shared and analysed to determine the appropriate reference for each student's performance. With this interactive process, teachers can establish a contextual performance image of what students know and can do along the developmental trajectory of creativity. A deeper understanding of the measurement rubric for creativity can boost teachers' confidence to use the generic measurement rubric or build an aligned task-specific rubric for scoring purposes in practice. When teachers can make a reliable judgement against the measurement rubric, student responses can be analysed quantitatively as other conventional examinations for statistical or comparative purposes as summative assessments. The same pool of student data allows for qualitative analysis and reporting, providing students with formative feedback on how well they had performed on a task and indicating their next level along the developmental continuum regarding creativity.

While this process is straightforward in terms of modern measurement theory, it differs from traditional creativity assessment, where scores are obtained by comparing students' responses with their peers in the same group.

8.5.2 Common Issues in Assessment Tasks for Creativity

On reviewing the initial draft of the tasks designed by the participating teachers in the second workshop, it was evident that two common issues occur when a teacher begins to conceptualise and develop a task for the measurement of creativity. The first issue concerns a restriction on the number of student responses that can be provided for a given task. An example of a task drafted to measure the component of problem finding is discussed below.

The initial task for problem finding asked students to discover three-to-five physics-based questions from a photo of people wearing masks on the street. Defining the number of responses required, in the stem of the task, was designed to encourage students to generate multiple answers, a departure from their usual approach to learning. If this is a common way of assessing, then there would be no need to limit the number of questions to three to five. Including such a requirement inhibits students who could come up with numerous inquiries relevant to their creative expression. In the literature, the proposal to limit the number of responses given by individuals was initially made to minimise the impact of the fluency index on the originality index in divergent-thinking tests (Michael & Wright, 1989). However, this assessment approach produces unnecessary interventions that impact creativity and

has been surpassed by more effective advancements in restructuring and computing the scoring metrics (Acar & Runco, 2012) and new developments in calculating semantic associative distance through Latent Semantic Analysis (Kenett, 2019; Volle, 2018). To promote the idea of generating multiple answers to a task, an alternative phrase derived from the Torrance's Test of Creative Thinking, "as many responses as possible", was considered by the participants as the best way to elicit multiple responses to nominated tasks. This draft task was therefore modified to ask students to discover as many physics-based questions as possible in response to the photo of people wearing masks on the street.

The second common issue that emerged in several of the participants' draft tasks was the lack of clarity in the creative characteristics of the expected outcome as per the construct of creativity. One question stemming from this topic is: What is the level of professional judgement exhibited by students who can produce numerous ideas versus those who can only come up with five ideas for a particular task? According to modern measurement theory, professional judgement in education is given by referencing students' performance to the established performance standards. While generating multiple responses is the intermediate process for inspiring and cultivating creativity, the construction of the performance standards in this thesis values the quality of response by demonstrating what students should know and can do in terms of creativity to differentiate between different levels of performance. The quality of response progresses from fluent thinking that allows for multiple responses from the same category to flexible thinking that produce different categories of

responses and then to a final creative response that is original and appropriate (see Section 3.3 on the process of building the performance standard for creativity for detail). Progression along the developmental continuum in creativity is related to the cognitive depth associated with the Revised Bloom's Taxonomy and should not be defined by the number of times a student demonstrates the same cognitive depth. It is important to feature the qualities of being novel and appropriate in the task description. This helps clarify the purpose of the assessment and motivates students to approach the task with a creative mindset. Therefore, the initial draft task for problem finding, above, was further modified to ask students to discover as many novel physics-based questions as possible that are appropriate to the photo of people wearing masks on the street.

The assessment of creativity in the literature is generally conducted through pen-and-paper formal tests, with the assessment requirements of novelty and appropriateness made clear in the administrative instructions for the test. On the other hand, the assessment tasks created through this research can be adapted to different teaching situations and presented to students in various formats. In this case, each task should be well designed to clarify what is being assessed and what a desired response might look like. To ensure that a task enables students to demonstrate the construct (creativity) being assessed, reflecting on "how I (the student) would perform when given a task" from the student's perspective can provide valuable insight into validating whether the task requires students to use their creativity in their responses.

The process of designing assessment tasks and the outcome of tasks designed

by the Chinese physics teachers who participated in the two teacher development workshops suggest a qualified outcome. Teachers would be able to develop appropriate tasks that produce evidence to locate students within a generic measurement rubric for creativity, if tailored professional support is provided to help them overcome their reflective challenges (i.e., understanding the measurement rubric for creativity and the task design for each of the components in the rubric in this thesis). These workshops further provide an evidential case of validity to support that a generic measurement rubric for creativity can be used in domain-specific contexts (Year 10 physics in China in this case). The next section discusses the implications of this thesis.

8.6 Implications of the Thesis

This thesis showcases the feasibility of applying standards-referenced assessment in creativity by illustrating the process of building a measurement rubric to describe, through the performance standards linked directly to the components of creativity, what students know and can do at each level of performance. Evidence is collected from qualitative content analysis of participants' comments and feedback on the extent to which the measurement rubric enables the monitoring of creativity across disciplinary contexts in high school settings. Additional evidence was gathered through two teacher development workshops, during which participants received professional support and shared their own experiences in creating assessment tasks aligned with the measurement rubric.

For a long time, assessments of creativity have typically involved comparing group performance through normative measurement (Fishkin & Johnson, 1998; Plucker & Makel, 2010; Said Metwaly et al., 2017). Such norm referencing does not provide standards for defining what growth looks like in creativity. Therefore, teachers in classrooms are unable to use using different assessments to assess and monitor student growth in creativity, or to compare student performance on different tasks. Building a measurement rubric for creativity is pioneering as it moves away from norm-referenced to a standards-referenced assessment model and develops performance standards linked to the components that evolve directly from the definition of creativity used in this thesis. These performance standards describe what students know and can do at each level of performance and consequently show a developmental continuum of creativity. The articulation of progress in creativity can effectively link assessment with the teaching and learning process. Teachers can know where individual students are and monitor their progress on the developmental continuum by referencing students' submitted responses to the tasks that the student attempts to the measurement rubric. This allows them to provide appropriate guidance to facilitate further learning and monitor longitudinal growth over time. This measurement of creativity also benefits students because they can get immediate feedback relating to where they are located on the construct and what they must do to get to the next level. This feedback can come from the teacher or, because the learning has been made "visible" through the rubric, it can also come from themselves (self-assessment) or their peers (peer assessment).

The process of building and validating a measurement rubric for assessing creativity in this thesis can be generalised to any learning context. Creativity has been incorporated into the curriculum as the highest learning goal in domain-specific syllabi (e.g., the *Scheme on Senior High School's Curriculum in China*) and domain-generic learning continua (e.g., the learning continuum of critical and creative thinking in the Australian curriculum). In contrast to the standardised learning trajectories of domain-specific knowledge and skills that have been recognised and documented in the curriculum, measurement rubrics that specifically demonstrate learning progress in creativity have not been clarified or determined. It can be challenging for teachers to fully understand the construct of creativity in teaching and assessment due to the absence of a clear definition and description of learning progression in student creativity. Yet, being exempt from high-stakes assessments empowers educators with the flexibility and liberty to define and measure what is truly valuable in terms of creativity for the 21st century. To prepare students for the ever-changing future, it is crucial to establish a learning continuum that focuses on generic and transferrable skills such as creativity. In fact, the process is generalisable and can be applied for use with other 21st-century skills. As an illustration, among the twelve essential skills for the 21st century, the "4Cs" stand out as the most significant – critical thinking, creativity, collaboration, and communication (Anderson & Jefferson, 2018). Numerous professional supports have been offered to facilitate transformative pedagogies and learning for the 4Cs (Anderson & Jefferson, 2021; Thornhill-Miller et al., 2023). Following the construction process outlined in this

thesis, educators at all levels of schooling can thoroughly examine various definitions and descriptions from the literature to develop a measurement rubric that suits their educational context and meets specific measurement objectives, such as creativity or the holistic 4Cs, for transformational education.

This thesis details how empirical evidence in the form of feedback from teachers and their submitted assessment tasks can be used to produce further evidence as to the validity of using a measurement rubric in classroom assessment. Hearing and engaging actively with teachers' voices is significant when building a learning continuum of creativity specific to an educational community, whether it be a class, a school or an education union. When teachers engage in professional conversations, they can share their knowledge and experience in teaching, learning and assessment. This helps to give practical insights into using a measurement rubric and ensuring the alignment between the intended measurement and the assessment result. Teachers wield the authority to employ the measurement rubric and assess their students' creativity. Nevertheless, when a rubric does not suit their teaching situation, it loses its effectiveness.

Participation in the design of measurement rubrics for creativity can also contribute to the professional development of teachers by providing an alternative perspective on assessment and measurement, and increasing their awareness of student creativity in teaching and learning. Involving teachers in designing appropriate tasks to provide evidence against a generic measurement rubric for creativity that can be reused is a new experience for Chinese physics teachers. The

main difference between the process of task design in this thesis and the normal process of using rubrics to mark student work is that the rubric remains constant and the tasks are altered. Generally, rubrics are written to capture the process the teacher would use to answer the question or task. Here the significant challenge is to design the task to elicit a response from the student that can be matched to the rubric. Besides, in traditional classroom assessments, defaulting to pen-and-paper methods within a discipline can sometimes limit teachers from understanding students' performance in creativity and exploring various assessment types. This can affect their ability to collect evidence of student performance and capture a developmental image in creativity according to the learning continuum in the measurement rubric. In a collaborative environment with teachers from different disciplinary backgrounds, teachers can immerse themselves in different assessment approaches and associated creative expression across disciplines. Teachers can develop their understanding of creativity in students and its measurement by interacting with student works and referencing them to the performance standards in the measurement rubric. When building a measurement rubric, involving teachers can be highly advantageous, particularly when the measurement takes into consideration the pedagogical foundation and classroom teaching. This allows for assessment evidence to inform customised instruction and improvement strategies, particularly in complex constructs such as creativity and other 21st-century skills.

8.7 Limitations and Future Directions

The measurement rubric for creativity was developed through a thorough review of existing research on various creativity theories and their assessment. Every development session was carefully reviewed and approved by the supervisory team, and the necessary modifications made during the validation process were communicated to the supervisors to ensure they accurately reflected the needs of the participating teachers. However, it would be beneficial if researchers in the field of creativity, particularly those who specialise in assessing creativity, were to participate in this study and review the rubric. Future studies on the measurement of creativity could conduct a panel review with experts in creativity to validate the definition of creativity, its components and the developmental continuum outlined in the performance standards. This approach can offer significant insights from a professional research standpoint.

Three focus groups were conducted as part of this research to validate the proposed measurement rubric for creativity. Due to the limited sample size, the measurement rubric resulting from the research only applies to teachers and researchers who participated in this research and readers who share the same understanding of creativity as defined in this thesis. If the measurement rubric is to be applied across a broader range of educational systems, additional focus groups and validation activities will be necessary.

Specifically, FG1 is a generic group comprising researchers and teachers from across learning disciplines to represent a broad view of disciplinary characteristics

and contexts for reviewing the measurement rubric. However, the disciplinary background of FG1 participants is not broad enough to include all the subjects taught in high school education. To better validate the generality of the measurement rubric for creativity, it is recommended to involve teachers from a wider field of disciplines in future research. FG2 and FG3 are specific groups in the context of science and physics, respectively, to validate the extent to which the generic measurement rubric can be used in this context. Including more disciplinary-specific group discussions is suggested to validate whether the generic measurement rubric can be used across disciplines. There are instances in which developing a measurement rubric specific for a particular domain may be required. Tailoring the generic rubric to specific contexts has the potential to reveal significant insights or characteristics for enhancing the generic measurement for creativity, which will be part of the validation process.

Three focus groups conducted in this research only lasted for two hours due to participants' availability, it may be beneficial to extend the focus group discussion to a duration of three hours or more to foster a more dynamic and engaging environment for a thorough and enlightening dialogue. The participants in FG3 were able to review the rubric for measuring creativity in a practical way by creating tasks related to different components of creativity and imagining how students would respond with reference to the performance-level descriptors. However, their limited knowledge of creativity resulted in a limitation for this study. It is recommended that participants with sufficient understanding of the construct (creativity in this case) be invited to participate in the validation process to provide more adequate input as evidence to

support the validity of the measurement rubric. To encourage a wider range of perspectives on student development and the alignment between the measurement rubric and the school setting, focus groups can involve not only teachers but also representatives from among policy makers, educational stakeholders and student representatives who have sufficient understanding of the construct of creativity. This will promote a diverse sharing of knowledge and experience that can be applied to building a dynamic schools and educational community that is energised beyond the immediate setting to promote student creativity as a consensual learning goal. The more effort that goes into validating the rubric and standardising the task, the greater the level of confidence that can be ascribed to the results from the process.

This thesis does not include the assessment practice of collecting student work for the assessment tasks and marking their performance by referencing them to the performance standards. The lack of assessment evidence makes it difficult to examine whether the designed assessment tasks can measure what they are intended to measure. Additionally, the consequential validity of using assessment results to inform teaching and learning in relation to creativity has not yet been evaluated. Incorporating students' actual task performance, alongside in-depth analysis, serves as a pivotal stride in using the measurement rubric within a real-world context for teacher performance development, particularly in task generation and 21st century skills instruction. Once students' task performance is collected and located on the measurement rubric, their performance level can be analysed using psychometric methods such as Rasch analysis or Item Response Theory to gain insights into the

quality of tasks and the growth of creativity in students.

The challenges reflected by the FGs participants, and the achievement made by the Chinese physics teachers in TDW provide some evidence about teacher growth with effective professional support, which may be a topic for future research. Due to the small sample size of this research, the teachers' challenges, the strategies and training program used cannot be generalised as protocols to promote teacher transformation in knowledge, thinking and practice. Future research could focus on teacher education in understanding and designing creativity assessments through analysing teachers' knowledge and beliefs, as well as developing a representative mechanism for teacher professional development (Calderhead, 1996; Clark & Peterson, 1986; Munby et al., 2001; Richardson, Placier, 2001; Woolfolk et al., 2006). This could be a significant achievement in filling the gap in assessment for creativity education in schools.

8.8 Concluding Remarks

The results of this research make possible the application of a standards-referenced model and modern measurement theory to develop a measurement rubric for the construct of creativity. In contrast to existing measures of creativity based on norm referencing, standards-referenced rubrics provide detailed descriptions of what progress and development in performance should look like, resulting in meaningful feedback on what it is the students know and can do in relation to creativity.

The thesis explicitly presents the process of building a measurement rubric for

creativity, which starts with defining the construct (creativity) being measured, followed by its measurable components and the associated performance standards that describe what students know and can do along the developmental continuum regarding creativity. The measurement rubric for creativity developed in this thesis was validated for its intended purpose, including its application across disciplines and in the specific domains of New South Wales science education and physics in the Chinese high school setting. Validation is constantly evolving, and it is possible for further exploration of the creativity assessment to provide additional arguments for validity and reliability beyond what has been researched to date.

It is believed that the process of development and validation of measurement rubrics has the potential to be generalised to other 21st-century constructs beyond creativity for the classroom, which has been previously considered challenging and demanding. Once the measurement rubric of a construct can be determined, it can equip teachers with a useful measurement tool for assessing and instructing their students effectively along its developmental continuum.

References

Acar, S., & Runco, M. A. (2012). Psychoticism and creativity: a meta-analytic review. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(4), 341-350. https://doi.org/10.1037/a0027497

Acar, S., & Runco, M. A. (2014). Assessing associative distance among ideas elicited by tests of divergent thinking. *Creativity Research Journal*, *26*(2), 229-238. https://doi.org/10.1080/10400419.2014.901095

ACARA. (2016a). *Critical and creative thinking (version 8.4)*. Australian Curriculum, Assessment and Reporting Authority. https://www.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/critical-and-creative-thinking/

ACARA. (2016b). *Critical and creative thinking learning continuum.* Australian Curriculum, Assessment and Reporting Authority. https://docs.acara.edu.au/resources/General_capabilities_-CCT_-_learning_continuum.pdf

ACARA. (2018). *NAP science literacy assessment framework*. Australian Curriculum, Assessment and Reporting Authority.

Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. *J Med Libr Assoc*, *103*(3), 152-153. https://doi.org/10.3163/1536-5050.103.3.010

Amabile, T. M. (1983a). A consensual technique for creativity assessment. In T. M. Amabile (Ed.), *The social psychology of creativity* (pp. 37-63). Springer New York. https://doi.org/10.1007/978-1-4612-5533-8_3

Amabile, T. M. (1983b). *The social psychology of creativity: a componential conceptualization*. American Psychological Association.

Amabile, T. M. (2018). *Creativity in context: update to the social psychology of creativity*. Routledge.

Amabile, T. M., Conti, R., Coon, H., Lazenby, J., & Herron, M. (1996). Assessing the work environment for creativity. *The Academy of Management Journal*, *39*(5), 1154-1184. https://doi.org/10.2307/256995

Amabile, T. M., & Pillemer, J. (2012). Perspectives on the social psychology of creativity. *The Journal of Creative Behavior*, *46*(1), 3-15. https://doi.org/https://doi.org/10.1002/jocb.001

American Educational Research Association. (2014). *Standards for educational and psychological testing (2014 edition)*. American Educational Research Association.

Ananiadoui, K., & Claro, M. (2009). 21st century skills and competences for new millennium learners in OECD countries. *OECD Education Working Papers*. https://doi.org/https://doi.org/10.1787/19939019

Anderson, M., & Jefferson, M. (2018). *Transforming organizations: engaging the 4cs for powerful organizational learning and change*. Bloomsbury Publishing.

Anderson, M., & Jefferson, M. (2021). *Transforming education: reimagining learning, pedagogy and curriculum*. Bloomsbury Publishing.

Andiliou, A., & Murphy, P. K. (2010). Examining variations among researchers' and teachers' conceptualizations of creativity: a review and synthesis of contemporary research. *Educational Research Review*, *5*(3), 201-219. https://doi.org/https://doi.org/10.1016/j.edurev.2010.07.003

Andrich, D., & Marais, I. (2019). *A course in Rasch measurement theory: measuring in the educational, social and health sciences*. Springer.

Angulo-Brunet, A., & Soto-Sanfiel, M. T. (2022). The link between creativity, personality, and art in communication students: Singapore. *Journalism & Mass Communication Educator*, *77*(3), 322-338. https://doi.org/10.1177/10776958211070462

Anitha, D., Kavitha, D., & Chandrasekaran, J. (2018). Assessing and enhancing creativity in a laboratory course with project based learning. *Journal of Engineering Education Transformations*, *32*. https://doi.org/10.16920/jeet/2018/v32i2/139505

Ayas, M. B., & Sak, U. (2014). Objective measure of scientific creativity: psychometric validity of the creative scientific ability test. *Thinking Skills and Creativity*, *13*, 195-205.

https://doi.org/https://doi.org/10.1016/j.tsc.2014.06.001

Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: the amusement park theoretical (APT) model of creativity. *Roeper Review*, *27*(3), 158-163. https://doi.org/10.1080/02783190509554310

Bapna, A., Sharma, N., Kaushik, A., & Kumar, A. (2017). *Handbook on measuring 21st century skills*. Evaldesign. https://doi.org/10.13140/RG.2.2.10020.99203

Barbot, B. (2019). Measuring creativity change and development. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 203-210. https://doi.org/https://doi.org/10.1037/aca0000232

Barbot, B., Besancon, M., & Lubart, T. (2011). Assessing creativity in the classroom. *The Open Education Journal*, *4*, 58-66. https://doi.org/10.2174/1874920801104010058

Barbot, B., Besançon, M., & Lubart, T. (2015). Creative potential in educational settings: its nature, measure, and nurture. *Education 3-13*, *43*(4), 371-381. https://doi.org/10.1080/03004279.2015.1020643

Barbot, B., Besançon, M., & Lubart, T. (2016). The generality-specificity of creativity: exploring the structure of creative potential with EPoC. *Learning and Individual Differences*, *52*, 178-187. https://doi.org/10.1016/j.lindif.2016.06.005

Barbot, B., Hass, R. W., & Reiter-Palmon, R. (2019). Creativity assessment in psychological research: (re)setting the standards. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 233-240. https://doi.org/10.1037/aca0000233

Barbot, B., & Reiter-Palmon, R. (2019). Creativity assessment: pitfalls, solutions, and standards. *Educational Publishing Foundation, 13.*

Batey, M., & Furnham, A. (2006). Creativity, intelligence, and personality: a critical review of the scattered literature. *Genetic, Social, and General Psychology Monographs*, *132*, 355-429. https://doi.org/10.3200/MONO.132.4.355-430

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: an open platform for computing semantic distance. *Behav Res Methods*, *53*(2), 757-780. https://doi.org/10.3758/s13428-020-01453-w

Beaty, R. E., Silvia, P. J., Nusbaum, E. C., Jauk, E., & Benedek, M. (2014). The roles of associative and executive processes in creative cognition. *Memory & Cognition*, *42*(7), 1186-1197. https://doi.org/10.3758/s13421-014-0428-8

Beghetto, R. A. (2013). Nurturing creativity in the micro-moments of the classroom. In K. H. Kim, J. C. Kaufman, J. Baer, & B. Sriraman (Eds.), *Creatively gifted students are not like other gifted students: research, theory, and practice* (pp. 3-15). SensePublishers. https://doi.org/10.1007/978-94-6209-149-8_2

Beghetto, R. A. (2018). *Beautiful risks: having the courage to teach and learn creatively*. Rowman & Littlefield.

Beghetto, R. A. (2019). Large-scale assessments, personalized learning, and creativity: paradoxes and possibilities. *ECNU Review of Education*, *2*(3), 311-327. https://doi.org/10.1177/2096531119878963

Beghetto, R. A., & Glăveanu, V. P. (2022). The beautiful risk of moving toward pedagogies of the possible. In *The Palgrave handbook of transformational giftedness for education* (pp. 23-42). Springer.

Beghetto, R. A., & Kaufman, J. C. (2014). Classroom contexts for creativity. *High Ability Studies*, *25*(1), 53-69. https://doi.org/https://doi.org/10.1080/13598139.2014.905247

Beghetto, R. A., & Sriraman, B. (2017). *Creative contradictions in education*. Springer.

Beketayev, K., & Runco, M. A. (2016). Scoring divergent thinking tests by computer with a semantics-based algorithm. *Eur J Psychol*, *12*(2), 210-220. https://doi.org/10.5964/ejop.v12i2.1127

Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., & Neubauer, A. C. (2017). How semantic memory structure and intelligence contribute to creative thought: a network science approach. *Thinking & Reasoning*, *23*(2), 158-183. https://doi.org/10.1080/13546783.2016.1278034

Bennett, J., Tognolini, J., & Pickering, S. (2012). Establishing and applying performance standards for curriculum-based examinations. *Assessment in Education: Principles, Policy & Practice*, *19*(3), 321-339.

https://doi.org/10.1080/0969594X.2011.614219

Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education*, *39*(1), 370-407. https://doi.org/10.3102/0091732x14554179

Bennett, R. E. (2018). Educational assessment: what to watch in a rapidly changing world. *Educational Measurement: Issues and Practice*, *37*(4), 7-15. https://doi.org/https://doi.org/10.1111/emip.12231

Bereczki, E. O., & Karpati, A. (2018). Teachers' beliefs about creativity and its nurture: a systematic review of the recent research literature. *Educational Research Review*, *23*, 25-56. https://doi.org/10.1016/j.edurev.2017.10.003

Bernadine, G. (2019). Challenges faced by educators in the implementation of Continuing Professional Teacher Development (CPTD): Gauteng Province. https://doi.org/10.5772/intechopen.84836

Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO Taxonomy. *Australian Journal of Education*, *33*(2), 151-163. https://doi.org/10.1177/168781408903300205

Biggs, J., & Collis, K. (2014). *Evaluating the quality of learning: the SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, *25*(6), 551-575. https://doi.org/10.1080/0969594X.2018.1441807

Black, P., Wilson, M., & Yao, S.-Y. (2011). Road maps for learning: a guide to the navigation of learning progressions. *Measurement: Interdisciplinary Research and Perspectives*, *9*(2-3), 71-123. https://doi.org/10.1080/15366367.2011.591654

Blamires, M., & Peterson, A. (2014). Can creativity be assessed? Towards an evidence-informed framework for assessing and planning progress in creativity. *Cambridge Journal of Education*, *44*(2), 147-162. https://doi.org/10.1080/0305764X.2013.860081

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956).

Taxonomy of educational objectives: the classification of educational goals. Handbook 1: cognitive domain. McKay New York.

Bolden, B., DeLuca, C., Kukkonen, T., Roy, S., & Wearing, J. (2020). Assessment of creativity in K-12 education: a scoping review. *Review of Education*, *8*(2), 343-376. https://doi.org/https://doi.org/10.1002/rev3.3188

Bond, T., Yan, Z., & Heene, M. (2020). *Applying the Rasch model: fundamental measurement in the human sciences*. Routledge.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences (2nd ed*.). Lawrence Erlbaum Associates Publishers.

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franić, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: revisions, new directions, and applications* (pp. 135–170). IAP Information Age Publishing.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061-1071. https://doi.org/10.1037/0033-295X.111.4.1061

Briggs, D. C., & Furtak, E. M. (2019). Learning progressions and embedded assessment. In *Classroom assessment and educational measurement* (pp. 146-169). Routledge.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5-12. https://doi.org/https://doi.org/10.1111/j.1745-3992.2003.tb00139.x

Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.

Brookhart, S. M. (2013a). Assessing creativity. *Educational Leadership*, *70*(5), 28-34.

Brookhart, S. M. (2013b). *How to create and use rubrics for formative assessment and grading*. Alexandria, Virginia USA : ASCD. https://search.library.wisc.edu/catalog/9910135062202121

Brookhart, S. M. (2017a). *How to give effective feedback to your students*. ASCD.

Brookhart, S. M. (2017b). *How to use grading to improve learning*. ASCD.

Brookhart, S. M. (2018). Appropriate criteria: key to effective rubrics. *Frontiers in Education*, *3*. https://doi.org/10.3389/feduc.2018.00022

Brookhart, S. M., & Chen, F. (2015). The quality and effectiveness of descriptive rubrics. *Educational Review*, *67*(3), 343-368. https://doi.org/10.1080/00131911.2014.929565

Brookhart, S. M., & McMillan, J. H. (2020). *Classroom assessment and educational measurement*. Taylor & Francis. https://doi.org/https://doi.org/10.4324/9780429507533

Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students*. Pearson Upper Saddle River, NJ.

Brown, G. (2017). *Assessment of student achievement*. Routledge. https://doi.org/10.4324/9781315162058

Brown, G., O'Leary, T., & Hattie, J. (2018). Effective reporting for formative assessment: the asTTle case example. In *Score reporting research and applications* (pp. 107-125). Routledge. https://doi.org/10.4324/9781351136501-9

Burnard, P., Colucci-Gray, L., & Cooke, C. (2022). Transdisciplinarity: re-visioning how sciences and arts together can enact democratizing creative educational experiences. *Review of Research in Education*, *46*(1), 166-197. https://doi.org/10.3102/0091732x221084323

Burnard, P., Colucci-Gray, L., & Sinha, P. (2021). Transdisciplinarity: letting arts and science teach together. *Curriculum Perspectives*, *41*(1), 113-118. https://doi.org/10.1007/s41297-020-00128-y

Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of Educational Psychology* (pp. 709-725). Simon & Schuster Macmillan.

Cambridge University Press. (n.d). Exploit. In *Cambridge dictionary*. Retrieved January 24, 2024 from

Care, E., Kim, H., Vista, A., & Anderson, K. (2019). Education system alignment for

21st century skills: focus on assessment. Brookings.

Care, E., Vista, A., & Kim, H. (2019). Assessment of transversal competencies: current tools in the Asian region. https://unesdoc.unesco.org/ark:/48223/pf0000368479.locale=en

Carson, S., Peterson, J., & Higgins, D. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal*, *17*, 37-50. https://doi.org/10.1207/s15326934crj1701_4

Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, *10*(6), 807-815. https://doi.org/https://doi.org/10.1016/j.cptl.2018.03.019

Chandio, M. T., Zafar, N., & Solangi, G. M. (2021). Bloom's Taxonomy: reforming pedagogy through assessment. *Journal of Education and Educational Development*, *8*(1), 109-140.

Chasteen, S. V., & Scherr, R. E. (2020). Developing the physics teacher education program analysis rubric: measuring features of thriving programs. *Physical Review Physics Education Research*, *16*(1), 010115. https://doi.org/10.1103/PhysRevPhysEducRes.16.010115

Childs, P., Han, J., Chen, L., Jiang, P., Wang, P., Park, D., . . . Vilanova, I. (2022). The Creativity Diamond-a framework to aid creativity. *J Intell*, *10*(4). https://doi.org/10.3390/jintelligence10040073

Chumo, C. C. (2014). Effects of practical investigation on scientific creativity amongst secondary schools biology students in Kericho district, Kenya. *Journal of Education and Practice*, *5*(8), 43-51.

Cizek, G. J. (2012). *Setting performance standards: foundations, methods, and innovations*. Routledge.

Cizek, G. J. (2020). *Validity: an integrated approach to test score meaning and use*. Routledge.

Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 255-296). MacMillan.

Claxton, G., & Lucas, B. (2013). What kind of teaching for what kind of learning.

SSAT.

Colangelo, N., Kerr, B., Hallowell, K., Huesman, R., & Gaeth, J. (1992). The Iowa inventiveness inventory: toward a measure of mechanical inventiveness. *Creativity Research Journal*, *5*(2), 157-163. https://doi.org/10.1080/10400419209534429

Core Literacy Research Group (2016). *Core literacy for student development in China*. Chinese Journal of Education.

Craft, A. (2006). Fostering creativity with wisdom. *Cambridge Journal of Education*, *36*(3), 337-350. https://doi.org/http://doi.org/10.1080/03057640600865835

Cremin, T., Burnard, P., & Craft, A. (2006). Pedagogy and possibility thinking in the early years. *Thinking Skills and Creativity*, *1*(2), 108-119.

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE Publications.

Cronbach, L. J. (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement (2nd ed.)* (pp. 443). American Council on Education.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281-302. https://doi.org/10.1037/h0040957

Cropley, D. H., Patston, T., Marrone, R. L., & Kaufman, J. C. (2019). Essential, unexceptional and universal: teacher implicit beliefs of creativity. *Thinking Skills and Creativity*, *34*, 100604. https://doi.org/https://doi.org/10.1016/j.tsc.2019.100604

Cseh, G., & Jeffries, K. (2019). A scattered CAT: A critical evaluation of the consensual assessment technique for creativity research. *Psychology of Aesthetics, Creativity, and the Arts*, *13*, 159-166. https://doi.org/10.1037/aca0000220

Csikszentmihalyi, M. (1996). The creative personality: flow and the psychology of discovery and invention. In *Creativity: The Psychology of Discovery and Invention* (pp. 480). HarperCollins.

Csikszentmihalyi, M., & Sawyer, K. (2015). Creative Insight: The Social Dimension of a Solitary Moment. In *The systems model of creativity* (pp. 73-98). Springer

Netherlands. https://doi.org/10.1007/978-94-017-9085-7_7

Dalby, D. (2019). Professional learning through collaborative research in mathematics. *Professional Development in Education*, *47*, 1-15. https://doi.org/10.1080/19415257.2019.1665571

Davis, G. A., & Rimm, S. (1982). Group Inventory for Finding Interests (GIFFI) I and II: instruments for identifying creative potential in the junior and senior high school. *The Journal of Creative Behavior*. https://doi.org/https://doi.org/10.1002/j.2162-6057.1982.tb00321.x

Dawson, P. (2015). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*. https://doi.org/10.1080/02602938.2015.1111294

Dede, C., Jass Ketelhut, D., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, *60*(1), 8-19. https://doi.org/10.1177/0022487108327554

Department for Education. (2014). *The national curriculum in England*. Retrieved from http://www.gov.uk/goverenment/publications/national-curriculum-in-england-secondary-curriculum

DeYoung, C. G. (2007). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*, 1138-1151. https://doi.org/10.1037/0022-3514.91.6.1138

DeYoung, C. G., Peterson, J. B., & Higgins, D. M. (2002). Higher-order factors of the Big Five predict conformity: are there neuroses of health? *Personality and Individual Differences*, *33*(4), 533-552. https://doi.org/https://doi.org/10.1016/S0191-8869(01)00171-4

Ding, M., & Chen, H. (2020). A study of the connotation of higher-order thinking skills in physics and its cognitive performance levels [master's thesis, Nanjing Normal University]. Nanjing Normal University.

Dollinger, S., Urban, K., & James, T. (2004). Creativity and openness: further validation of two creative product measures. *Creativity Research Journal*, *16*, 35-47. https://doi.org/10.1207/s15326934crj1601_4

Dreyfus, S. E., & Dreyfus, H. (1980). A Five-Stage model of the mental activities involved in directed skill acquisition. *Distribution*, 22.

DuFour, R., & DuFour, R. (2013). *Learning by doing: a handbook for professional learning communities at work TM*. Solution Tree Press.

DuFour, R., & Eaker, R. (2009). *Professional learning communities at work tm: best practices for enhancing students achievement*. Solution Tree Press.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107-115. https://doi.org/https://doi.org/10.1111/j.1365-2648.2007.04569.x

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior & adult)*. Hodder and Stoughton Educational.

Feist, G. J. (2010). The function of personality in creativity: the nature and nurture of the creative personality. In J. C. Kaufman & R. J. Sternberg (Eds.), *The Cambridge handbook of creativity* (pp. 113-130). Cambridge University Press. https://doi.org/https://doi.org/10.1017/CBO9780511763205.009

Feist, G. J. (2019). Creativity and the Big Two model of personality: plasticity and stability. *Current Opinion in Behavioral Sciences*, *27*, 31-35. https://doi.org/https://doi.org/10.1016/j.cobeha.2018.07.005

Feist, G. J., & Barron, F. X. (2003). Predicting creativity from early to late adulthood: Intellect, potential, and personality. *Journal of Research in Personality*, *37*(2), 62-88. https://doi.org/https://doi.org/10.1016/S0092-6566(02)00536-6

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: theory, research, and applications*. The MIT Press.

Finke, R. A., Ward, T. B., & Smith, S. M. (1996). *Creative cognition: theory, research, and applications (featured edition)*. The MIT press.

Fishkin, A. S., & Johnson, A. S. (1998). Who is creative? Identifying children's creative abilities. *Roeper Review*, *21*(1), 40-46. https://doi.org/10.1080/02783199809553925

Forster, M., & Masters, G. (2010). Progression and assessment: developmental assessment. In *International encyclopedia of education* (3rd ed.). Elsevier

Science. https://doi.org/10.1016/B978-0-08-044894-7.00354-7

Freire, P. (1982). Creating alternative research methods: Learning to do it by doing it. In B. Hall, A Gillette, & R. Tandon (Eds.), *Creating knowledge: a monopoly? Participatory research in development* (pp. 29-37). New Delhi, India: Participatory Research in Asia.

Fullan, M. (2007). *The new meaning of educational change (4th ed.)*. Teachers College Press.

Getzels, J. W., & Csikszentmihalyi, M. (1976). The creative vision: a longitudinal study of problem finding in art. *Wiley*.

Glăveanu, V. P. (2018). Educating which creativity? *Thinking Skills and Creativity*, *27*, 25-32. https://doi.org/https://doi.org/10.1016/j.tsc.2017.11.006

Glăveanu, V. P. (2020). A sociocultural theory of creativity: bridging the social, the material, and the psychological. *Review of General Psychology*, *24*, 335-354. https://doi.org/10.1177/1089268020961763

Glaveanu, V. P., Hanchett Hanson, M., Baer, J., Barbot, B., Clapp, E. P., Corazza, G. E., . . . Sternberg, R. J. (2020). Advancing creativity theory and research: a socio-cultural manifesto. *The Journal of Creative Behavior*, *54*(3), 741-745. https://doi.org/https://doi.org/10.1002/jocb.395

Goldberg, L. R. (2013). An alternative “description of personality”: the Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216-1129. https://doi.org/https://doi.org/10.1037/0022-3514.59.6.1216

Gordon, W. J., & Poze, T. (1980). *The new art of the possible: the basic course in synectics*. Trackmaster Porpoise Books.

Grbich, C. (2007). *Qualitative data analysis: an introduction*. SAGE Publications.

Guilford, J. P. (1950). Creativity. *The American Psychologist*, *5*(9), 444-454. https://doi.org/https://doi.org/10.1037/h0063487

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*(4), 267-293. https://doi.org/10.1037/h0040755

Guilford, J. P. (1959). Three faces of intellect. *The American Psychologist*, *14*(8), 469. https://doi.org/https://doi.org/10.1037/h0046827

Guilford, J. P. (1966). Measurement and creativity. *Theory Into Practice*, *5*(4), 185-189. https://doi.org/https://doi.org/10.1080/00405846609542023

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, *25*(12), 1509-1528. https://doi.org/10.1080/0950069022000038268

Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items.* Routledge.

Hammer, S. (2007). Demonstrating quality outcomes in learning and teaching: examining 'best practice' in the use of criterion-referenced assessment. *International Journal of Pedagogies and Learning*, *3*(1), 50-58. https://doi.org/10.5172/ijpl.3.1.50

Hanif, S., Wijaya, A. F. C., & Winarno, N. (2019). Enhancing students' creativity through stem project-based learning. *Journal of Science Learning*, *2*(2), 50-57.

Harris, D., & Bruin, L. (2017). Training teachers for twenty-first century creative and critical thinking: Australian implications from an international study. *Teaching Education*, *29*, 1-17. https://doi.org/10.1080/10476210.2017.1384802

Hass, R. W. (2017). Tracking the dynamics of divergent thinking via semantic distance: analytic methods and theoretical implications. *Memory & Cognition*, *45*(2), 233-244. https://doi.org/10.3758/s13421-016-0659-y

Hendriks, A. J., Hofstee, W. K., & De Raad, B. (1999). The five-factor personality inventory (FFPI). *Personality and Individual Differences*, *27*(2), 307-325. https://doi.org/https://doi.org/10.1016/S0191-8869(98)00245-1

Hendry, G. D., Armstrong, S., & Bromberger, N. (2012). Implementing standards-based assessment effectively: incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, *37*(2), 149-161. https://doi.org/10.1080/02602938.2010.515014

Hennink, M. M. (2014). *Focus group discussions*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199856169.001.0001

Henriksen, D., & Mishra, P. (2015). We teach who we are: creativity in the lives and

practices of accomplished teachers. *Teachers College Record*, *117*, 1-46. https://doi.org/10.1177/016146811511700708

Henriksen, D., Mishra, P., & the Deep-Play Research, G. (2020). A pragmatic but hopeful conception of creativity: a conversation with Dr. Barbara Kerr. *TechTrends*, *64*(2), 195-201. https://doi.org/10.1007/s11528-020-00476-6

Hocevar, D. (1979). The development of the creative behavior inventory (CBI).

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science Education*, *24*(4), 389-403. https://doi.org/10.1080/09500690110098912

Huang, P.-S., Peng, S.-L., Chen, H.-C., Tseng, L.-C., & Hsu, L.-C. (2017). The relative influences of domain knowledge and domain-general divergent thinking on scientific creativity and mathematical creativity. *Thinking Skills and Creativity*, *25*, 1-9. https://doi.org/https://doi.org/10.1016/j.tsc.2017.06.001

Isaksen, S. G., & Treffinger, D. J. (1985). *Creative problem solving: the basic course.* New York: Bearly Limited.

John, O., Naumann, L., & Soto, C. (2008). Paradigm shift to the integrative big five trait taxonomy: History, measurement, and conceptual issues. In O. John, R. W. Robins & L. A. Pervin (Eds.), *Handbook of personality: theory and research* (3rd ed., pp. 114-158). Guilford Press.

Jonsson, A. (2014). Rubrics as a way of providing transparency in assessment. *Assessment & Evaluation in Higher Education*, *39*(7), 840-852. https://doi.org/10.1080/02602938.2013.875117

Jönsson, A., & Panadero, E. (2016). The use and design of rubrics to support assessment for learning. In D. Carless, S. Bridges, C. Chan, R. Glofcheski (Eds.), *Scaling up assessment for learning in higher education* (pp. 99-111). Springer Singapore. https://doi.org/10.1007/978-981-10-3045-1_7

Jönsson, A., & Svingby, G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, *2*, 130-144. https://doi.org/10.1016/j.edurev.2007.05.002

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (Vol. 3). Association for Supervision and Curriculum Development. Alexandria, VA.

Kamei, T. (2019). *Thinking to learn: an investigation of assessing and teaching thinking skills for students with disability* [Doctoral dissertation, University of Melbourne]. University of Melbourne.

Kamei, T. (2020). Thinking skills instructional strategies: teaching students with additional needs to be better thinkers. In Griffin, P., Woods, K. (Ed.), *Understanding students with additional needs as learners* (pp. 185-195). Springer Cham. https://doi.org/10.1007/978-3-030-56596-1_12

Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, *64*(3), 425-461. https://doi.org/10.2307/1170678

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). American Council on Education/Praeger.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448-457.

Kane, M. (2016). Explicating validity. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 198-211. https://doi.org/10.1080/0969594X.2015.1060192

Karwowski, M. (2021). School does not kill creativity. *European Psychologist*, *27*. https://doi.org/10.1027/1016-9040/a000449

Kaufman, J., C, Kapoor, H., Patston, T., & Cropley, D. (2021). Explaining standardized educational test scores: the role of creativity above and beyond GPA and personality. *Psychology of Aesthetics, Creativity, and the Arts*. https://doi.org/10.1037/aca0000433

Kaufman, J. C., & Baer, J. (2005). *Creativity across domains: faces of the muse*. Lawrence Erlbaum Associates Publishers.

https://doi.org/10.4324/9781410611925

Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: the Four C model of creativity. *Review of General Psychology*, *13*(1), 1-12. https://doi.org/10.1037/a0013688

Kaufman, J. C., & Beghetto, R. A. (2013). Do people recognize the four Cs? Examining layperson conceptions of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 229-236. https://doi.org/10.1037/a0033295

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. John Wiley & Sons.

Kaufman, J. C., & Sternberg, R. J. (2019). *The cambridge handbook of creativity* (2 ed.). Cambridge University Press. https://doi.org/DOI: 10.1017/9781316979839

Kaufman, S. B., Quilty, L. C., Grazioplene, R. G., Hirsh, J. B., Gray, J. R., Peterson, J. B., & DeYoung, C. G. (2016). Openness to experience and intellect differentially predict creative achievement in the arts and sciences. *Journal of Personality*, *84*(2), 248-258. https://doi.org/https://doi.org/10.1111/jopy.12156

Kenett, Y. N. (2018). Investigating creativity from a semantic network perspective. In Z. Kapoula, E. Volle, J. Renoult, M. Andreatta (Eds.), *Exploring trandisciplinarity in art and sciences* (pp. 49-75). Spinger Cham. https://doi.org/10.1007/978-3-319-76054-4_3

Kenett, Y. N. (2019). What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, *27*, 11-16. https://doi.org/https://doi.org/10.1016/j.cobeha.2018.08.010

Kenett, Y. N., Anaki, D., & Faust, M. (2014). Investigating the structure of semantic networks in low and high creative persons. *Front Hum Neurosci*, *8*, 407. https://doi.org/10.3389/fnhum.2014.00407

Kenett, Y. N., Gooz, N., & Ackerman, R. (2023). The role of semantic associations as a metacognitive cue in creative idea generation. *J Intell*, *11*(4). https://doi.org/10.3390/jintelligence11040059

Kettler, T., Lamb, K. N., Willerson, A., & Mullet, D. R. (2018). Teachers' perceptions

of creativity in the classroom. *Creativity Research Journal*, *30*(2), 164-171. https://doi.org/https://doi.org/10.1080/10400419.2018.1446503

Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity Research Journal*, *18*(1), 3-14. https://doi.org/10.1207/s15326934crj1801_2

Kim, K. H. (2011). The creativity crisis: the decrease in creative thinking scores on the Torrance Tests of Creative Thinking. *Creativity Research Journal*, *23*(4), 285-295. https://doi.org/https://doi.org/10.1080/10400419.2011.627805

Klahr, D. (2003). Review of exploring science: the cognition and development of discovery processes. *Human Development*, *46*(2-3), 155-160. https://doi.org/10.1159/000068588

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, *12*(1), 1-48. https://doi.org/10.1207/s15516709cog1201_1

Krathwohl, D. R. (2002). A revision of Bloom's Taxonomy: an overview. *Theory Into Practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Lamri, J., & Lubart, T. (2021). Creativity and its' relationships with 21st century skills in job performance. *Kindai Management Review*, *9*, 75-91.

Lane, S. (2013). *SAGE handbook of research on classroom assessment.* SAGE Publications. https://doi.org/10.4135/9781452218649

Lane, S., & Tierney, S. T. (2008). 21st century education: a reference handbook. SAGE Publications. https://doi.org/10.4135/9781412964012

Li, T., He, P., & Peng, L. (2023). Measuring high school student engagement in science learning: an adaptation and validation study. *International Journal of Science Education*. https://doi.org/10.1080/09500693.2023.2248668

Lorge, I. D. (1951). The fundamental nature of measurement. In E. F. Lindquist (Ed.), *Educational measurement* (pp. 533-559). American Council on Education.

Lubart, T., Besançon, M., & Barbot, B. (2011). *EPOC: évaluation du potentiel créatif*. Hogrefe.

Lubart, T., Pacteau, C., Jacquet, A.-Y., & Caroff, X. (2010). Children's creative potential: an empirical study of measurement issues. *Learning and Individual*

Differences, *20*(4), 388-392. https://doi.org/https://doi.org/10.1016/j.lindif.2010.02.006

Lubart, T., Zenasni, F., & Barbot, B. (2013). Creative potential and its measurement. *International Journal for Talent Development and Creativity*, *1*, 41-50.

Lubart, T. B., Baptiste; Besancon, Maud. (2019). Creative potential: assessment issues and the EPoC Battery. *Estudios de Psicología*, *40*, 1-23. https://doi.org/10.1080/02109395.2019.1656462

Lucas, B. (2016). A Five-Dimensional model of creativity and its assessment in schools. *Applied Measurement in Education*, *29*(4), 278-290. https://doi.org/10.1080/08957347.2016.1209206

Lucas, B. (2019). Teaching and assessing creativity in schools in England. *Imapct Journal*(7), 4-8.

Lucas, B. (2022). *A field guide to assessing creative thinking in schools*. FORM. https://doi.org/10.13140/RG.2.2.24010.03529

Lucas, B., Claxton, G., & Spencer, E. (2013). *Progression in student creativity in school: first steps towards new forms of formative assessments*. OECD. https://doi.org/https://doi.org/10.1787/19939019

Lucas, B., Spencer, E., Stoll, L., Fisher-Naylor, D., Richards, N., James, S., & Milne, K. (2023). *Creative thinking in schools: a leadership playbook*. Crown House Publishing Ltd.

MacCann, R. G., & Stanley, G. (2010). Extending participation in standard setting: an online judging proposal. *Educational Assessment, Evaluation and Accountability*, *22*(2), 139-157. https://doi.org/10.1007/s11092-010-9094-y

Mangion, M., & Riebel, J. A. (2023). Young creators: perceptions of creativity by primary school students in Malta. *J Intell*, *11*(3). https://doi.org/10.3390/jintelligence11030053

Margolis, M. J., Mee, J., Clauser, B. E., Winward, M., & Clauser, J. C. (2016). Effect of content knowledge on Angoff-Style standard setting judgements. *Educational Measurement: Issues and Practice*, *35*(1), 29-37. https://doi.org/https://doi.org/10.1111/emip.12104

Masia, B. B., & Bloom, B. S. (1964). *Taxonomy of educational objectives: the classification of educational goals: handbook II: affective domain.* David McKay Company, Inc.

Masters, G. N. (2013). Reforming educational assessment: imperatives, principles and challenges. *Australian Education Review, 57*. Retrieved from http://research.acer.edu.au/aer/12/

Masters, G. N., & Forster, M. (1996). Progress maps: assessment resource kit. *Australian Council for Educational Research.*

Mayring, P. (2004). Qualitative content analysis. In Uwe Flick, Ernst von Kardoff, Ines Steinke (Eds.), *A companion to qualitative research* (pp, 159-176). SAGE Publications.

Mayring, P. (2015). Qualitative content analysis: theoretical background and procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: examples of methodology and methods* (pp. 365-380). Springer Netherlands. https://doi.org/10.1007/978-94-017-9181-6_13

McClure, E., & Jaeger, G. (2020). Assessing creativity: a palette of possibilities. *Billund: LEGO Foundation*.

McKenna, M., Dedrick, R. F., & Goldstein, H. (2022). Development and initial validation of the early elementary writing rubric to inform instruction for kindergarten and first-grade students. *Assessment for Effective Intervention*, *47*(4), 220-233. https://doi.org/10.1177/15345084211065977

McMillan, J. (2013). *SAGE handbook of research on classroom assessment.* SAGE Publications. https://doi.org/10.4135/9781452218649

McMillan, J. H. (2014). *Classroom assessment: principles and practice for effective standards-based instruction*. Pearson.

Mednick, S. (1962). *The associative basis of the creative process.* American Psychological Association. http://doi.org/10.1037/h0048850

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: a guide to design and implementation*. John Wiley & Sons.

Mertler, C. (2001). Designing scoring rubrics for your classroom. *Research & Evaluation*, *7*.

Merton, R., Fiske, M., & Kendall, P. (1956). *The focused interview: a manual of problems and procedures.* Free Press.

Messick, S. (1989a). Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. https://doi.org/10.3102/0013189X018002005

Messick, S. (1989b). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). American Council on Education.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*(2), 13-23. https://doi.org/10.3102/0013189X023002013

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50, 741-749*. https://doi.org/10.1037/0003-066X.50.9.741

Messick, S. (1998). Consequences of test interpretation and use: the fusion of validity and values in psychological assessment. *ETS Research Report Series*, *1998*(2), i-32. https://doi.org/https://doi.org/10.1002/j.2333-8504.1998.tb01797.x

Michael, W. B., & Wright, C. R. (1989). Psychometric issues in the assessment of creativity. In *Handbook of creativity* (pp. 33-52). Springer.

Miller, D. M., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and assessment in teaching (11th ed.)*. Pearson.

Ministry of Education of People's Republic of China. (2017a). *Scheme on senior high school's curriculum*. People's Education Press.

Ministry of Education of People's Republic of China. (2017b). *Standard of the physics curriculum for senior high schools.* People's Education Press.

Mishra, L. (2016). Focus group discussion in qualitative research. *TechnoLearn, 6*, 1. https://doi.org/10.5958/2249-5223.2016.00001.2

Mishra, P., & Henriksen, D. (2013). A new approach to defining and measuring

creativity: rethinking technology & creativity in the 21st century. *TechTrends*, *57*(5), 10-13. https://doi.org/https://doi.org/10.1007/s11528-013-0685-6

Morgan, D. L. (1996). *Focus groups as qualitative research* (Vol. 16). SAGE Publications.

Morgan, D. L., & Hoffman, K. (2018). *The SAGE handbook of qualitative data collection.* SAGE Publications. https://doi.org/10.4135/9781526416070

Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assessment, Research, and Evaluation*, *7*(1), 3.

Mumford, M. D., Giorgini, V., Gibson, C., & Mecca, J. (2013). Creative thinking: processes, strategies and knowledge. In *Handbook of research on creativity* (pp. 249-264). Edward Elgar Publishing. https://doi.org/10.4337/9780857939814

Mumford, M. D., & McIntosh, T. (2017). Creative thinking processes: the past and the future. *The Journal of Creative Behavior*, *51*(4), 317-322. https://doi.org/https://doi.org/10.1002/jocb.197

Mumford, M. D., Medeiros, K. E., & Partlow, P. J. (2012). Creative thinking: processes, strategies, and knowledge. *The Journal of Creative Behavior*, *46*(1), 30-47. https://doi.org/https://doi.org/10.1002/jocb.003

Munby, H., Russell, T., & Martin, A. K. (2001). Teachers' knowledge and how it develops. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th ed., pp. 877-904). AERA.

Myszkowski, N., & Storme, M. (2019). Judge response theory? A call to upgrade our psychometrical account of creativity judgements. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(2), 167-175. https://doi.org/10.1037/aca0000225

NACCCE. (1999). *All our futures: creativity, culture and education*. https://sirkenrobinson.com/pdf/allourfutures.pdf

National Education Association of the United States. (2009). 21st Century Skills, education and competitiveness: a resource and policy guide. *Partnership for 21st Century Skills*, *21*.

NESA. (2021). *NSW curriculum and syllabuses*. NSW Education Standards Authority

Retrieved from https://educationstandards.nsw.edu.au/wps/portal/nesa/11-12/Understanding-the-curriculum/nsw-curriculum-syllabuses

Neuendorf, K. A. (2018). Content analysis and thematic analysis. In *Advanced research methods for applied psychology: design, analysis and reporting* (pp. 211-305). Routledge.

Newton, P. E. (2012). Questioning the consensus definition of validity. *Measurement: Interdisciplinary Research and Perspectives*, *10*(1-2), 110-122. https://doi.org/10.1080/15366367.2012.688456

Nitko, A. J. (1980). Criterion-referencing schemes. *New directions for testing and measurement*, *6*, 35-71.

Nitko, A. J. (2001). *Educational assessment of students (3rd ed.)*. Merrill.

Nuhoglu, H., & Akgül, S. (2019). Analysis of the relation between creativity level and problem solving skills of gifted and talented students. *Educational Research and Reviews*, *14*(15), 518-532.

OECD. (2018a). *Preparing our youth for an inclusive and sustainable world: the OECD PISA global competence framework*. OECD. https://www.oecd.org/pisa/Handbook-PISA-2018-Global-Competence.pdf

OECD. (2018b). *Questionns related to global competence in the student questionnaire*. OECD. https://www.oecd.org/pisa/PISA-2018-Global-Competence-Questionnaire.pdf

OECD. (2019). PISA 2021 creative thinking framework (3rd draft). https://www.oecd.org/pisa/publications/PISA-2021-creative-thinking-framework.pdf

OECD. (2023a). *Education at a glance 2023*. https://doi.org/doi:https://doi.org/10.1787/eag-2018-en

OECD. (2023b). *PISA 2022 results (Volumne I): the state of learning and equity in education*. OECD.

Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2009). A qualitative framework for collecting and analyzing data in focus group research. *International Journal of Qualitative Methods*, *8*(3), 1-21.

https://doi.org/10.1177/160940690900800301

Ozkan, G., & Umdu Topsakal, U. (2021). Exploring the effectiveness of STEAM design processes on middle school students' creativity. *International Journal of Technology and Design Education*, *31*(1), 95-116. https://doi.org/10.1007/s10798-019-09547-z

Panadero, E., & Jonsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, *30*, 100329. https://doi.org/https://doi.org/10.1016/j.edurev.2020.100329

Parnes, S. J. (1981). *The magic of your mind*. Creative Education Foundation.

Patston, T. J., Kaufman, J. C., Cropley, A. J., & Marrone, R. (2021). What is creativity in education? A qualitative study of international curricula. *Journal of Advanced Academics*, *32*(2), 207-230. https://doi.org/https://doi.org/10.1177/1932202X209783

Perignat, E., & Katz-Buonincontro, J. (2019). STEAM in practice and research: an integrative literature review. *Thinking Skills and Creativity*, *31*, 31-43. https://doi.org/https://doi.org/10.1016/j.tsc.2018.10.002

Perkins, D. N., & Perkins, D. N. (2009). *The mind's best work*. Harvard University Press.

Plucker, J. A. (1999). Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal data. *Creativity Research Journal*, *12*(2), 103-114. https://doi.org/10.1207/s15326934crj1202_3

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist*, *39*(2), 83-96. https://doi.org/10.1207/s15326985ep3902_1

Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In *The Cambridge handbook of creativity* (pp. 48-73). Cambridge University Press. https://doi.org/10.1017/CBO9780511763205.005

Popham, W. J. (1971). *Criterion-referenced measurement: an introduction*. Educational Technology.

Powell, C. G., & Bodur, Y. (2019). Teachers' perceptions of an online professional development experience: implications for a design and implementation framework. *Teaching and Teacher Education*, *77*, 19-30.

Prince, G. M. (1968). The operational mechanism of synectics. *The Journal of Creative Behavior*, *2*(1), 1-13. https://doi.org/https://doi.org/10.1002/j.2162-6057.1968.tb00077.x

Puryear, J. S., & Lamb, K. N. (2020). Defining creativity: how far have we come since Plucker, Beghetto, and Dow? *Creativity Research Journal*, *32*(3), 206-214. https://doi.org/10.1080/10400419.2020.1821552

QCA. (2005). *Creativity: find it, promote it!—promoting pupils' creative thinking and behaviour across the curriculum at key stages 1, 2 and 3—practical materials for schools.* Qualifications and Curriculum Authority.

QCA. (2009). *A framework of personal, learning and thinking skills 11-19 in England.* Qualifications and Curriculum Authority. https://dera.ioe.ac.uk/id/eprint/7268/3/PLTS_framework_v2_tcm8-936-1.pdf

Qian, M., Plucker, J. A., & Yang, X. (2019). Is creativity domain specific or domain general? Evidence from multilevel explanatory item response theory models. *Thinking Skills and Creativity*, *33*, 100571. https://doi.org/https://doi.org/10.1016/j.tsc.2019.100571

Reiter-Palmon, R. (2020). The consensual assessment technique: refinement and further development. In R. Reiter-Palmon, M. Fisher, J. S. Mueller (Eds.), *Creativity at work* (pp. 157-165). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-61311-2_15

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*(7), 305-310. http://www.jstor.org/stable/20342603

Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th ed., pp. 905-947). AERA.

Robinson, K., & Lee, J. R. (2011). *Out of our minds*. Wiley Online Library.

Rodriguez, R., Silvia, P., Kaufman, J., Reiter-Palmon, R., & Puryear, J. (2023). Taking inventory of the creative behavior inventory: an Item Response Theory

analysis of the CBI. *Creativity Research Journal*, 1-11. https://doi.org/10.1080/10400419.2023.2183322

Rohr-Mentele, S., & Forster-Heinzer, S. (2021). Practical validation framework for competence measurement in VET: a validation study of an instrument for measuring basic commercial knowledge and skills in Switzerland. *Empirical Research in Vocational Education and Training*, *13*. https://doi.org/10.1186/s40461-021-00122-2

Runco, M. A. (1991). *Divergent thinking*. Ablex Publishing.

Runco, M. A. (2014). *Creativity: Theories and Themes: Research, Development, and Practice*. Elsevier. https://doi.org/10.1016/C2012-0-06920-7

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal*, *24*(1), 66-75. https://doi.org/10.1080/10400419.2012.652929

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92-96. https://doi.org/10.1080/10400419.2012.650092

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, *13*, 191-209.

Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, *30*, 175-194. https://doi.org/10.1080/0260293042000264262

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, *34*(2), 159-179. https://doi.org/10.1080/02602930801956059

Sahlberg, P. (2009). The role of education in promoting creativity: potential barriers and enabling factors. *Measuring creativity*, 337.

Said Metwaly, S., Van den Noortgate, W., & Kyndt, E. (2017). Approaches to measuring creativity: a systematic literature review. *Creativity Theories – Research – Applications*, *4*. https://doi.org/10.1515/ctra-2017-0013

Sak, U., & Ayas, M. B. (2013). Creative scientific ability test (C-SAT): a new measure

of scientific creativity. *Psychological Test and Assessment Modeling*, *55*, 328.

Sariscsany, M. J. (1996). Turning ideas into practices: a teacher development workshop. *Journal of Physical Education, Recreation & Dance*, *67*(3), 38-40. https://doi.org/10.1080/07303084.1996.10607219

Scalise, K., & Wilson, M. (2011). The nature of assessment systems to support effective use of evidence through technology. *E-Learning and Digital Media*, *8*(2), 121-132. https://doi.org/10.2304/elea.2011.8.2.121

Scates, D. E. (1943). Differences between measurement criteria of pure scientists and of classroom teachers. *The Journal of Educational Research*, *37*(1), 1-13. http://www.jstor.org/stable/27528418

Schleicher, A. (2019). *PISA 2018: insights and interpretations*. OECD. Retrieve from https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf

Shaw, A. (2022). Creative Minecrafters: cognitive and personality determinants of creativity, novelty, and usefulness in Minecraft. *Psychology of Aesthetics Creativity and the Arts*, *17*. https://doi.org/10.1037/aca0000456

Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623-646). Praeger.

Shepard, L. A. (2016). Evaluating test validity: reprise and progress. *Assessment in Education: Principles, Policy & Practice*, *23*(2), 268-280. https://doi.org/10.1080/0969594X.2016.1141168

Shepard, L. A. (2019). Classroom assessment to support teaching and learning. *The ANNALS of the American Academy of Political and Social Science*, *683*(1), 183-200. https://doi.org/10.1177/0002716219843818

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21-34. https://doi.org/https://doi.org/10.1111/emip.12189

Simonton, D. K. (1980). Thematic fame, melodic originality, and musical zeitgeist: a biographical and transhistorical content analysis. *Journal of Personality and*

Social Psychology, *38*(6), 972-983. https://doi.org/https://doi.org/10.1037/0022-3514.38.6.972

Simonton, D. K. (1998). Creativity from a historiometric perspective. In R. J. Sternberg (Ed.), *Handbook of Creativity* (pp. 116-134). Cambridge University Press. https://doi.org/10.1017/CBO9780511807916.008

Skiba, T., Tan, M., Sternberg, R. J., & Grigorenko, E. L. (2010). Roads not taken, new roads to take: looking for creativity in the classroom. In *Nurturing creativity in the classroom* (pp. 252-269). Cambridge University Press. https://doi.org/10.1017/CBO9780511781629.013

Sparkes, A. C. (2005). Narrative analysis: exploring the whats and hows of personal stories. *Qualitative Research in Health Care*, *1*(1), 191-209.

Starko, A. (2013). *Creativity in the classroom: schools of curious delight*. Routledge.

Sternberg, R. J. (1986). Intelligence, wisdom, and creativity: three is better than one. *Educational Psychologist*, *21*(3), 175-190. https://doi.org/10.1207/s15326985ep2103_2

Sternberg, R. J. (1996). *Successful intelligence : how practical and creative intelligence determine success in life*. Simon & Schuster.

Sternberg, R. J. (2018). A triangular theory of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *12*(1), 50-67. https://doi.org/10.1037/aca0000095

Sternberg, R. J. (2015). Teaching for creativity: the sounds of silence. *Psychology of Aesthetics, Creativity, and the Arts*, *9*. https://doi.org/10.1037/aca0000007

Sternberg, R. J., & Lubart, T. I. (1992a). Buy low and sell high: an investment approach to creativity. *Current Directions in Psychological Science*, *1*(1), 1-5. https://doi.org/10.1111/1467-8721.ep10767737

Sternberg, R. J., & Lubart, T. I. (1992b). Creativity: its nature and assessment. *School Psychology International*, *13*(3), 243-253. https://doi.org/10.1177/0143034392133004

Sternberg, R. J., & Lubart, T. I. (1995). *Defying the crowd: cultivating creativity in a culture of conformity.* Free Press.

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: prospects and

paradigms. In *Handbook of creativity* (pp. 3-15). Cambridge University Press.

Steyn, G. M., & Niekerk, L. J. V. (2005). Professional development of teachers : critical success factors. *Koers : Bulletin for Christian Scholarship*, *70*(1), 125-149. https://doi.org/doi:10.10520/EJC59178

Storer, T. (2018). *The effect of project based learning on the creativity of elementary students*. Wilkes University.

Sun, M., Wang, M., & Wegerif, R. (2019). Using computer-based cognitive mapping to improve students' divergent thinking for creativity development. *British Journal of Educational Technology*, *50*(5), 2217-2233. https://doi.org/https://doi.org/10.1111/bjet.12825

Sun, M., Wang, M., & Wegerif, R. (2020). Effects of divergent thinking training on students' scientific creativity: the impact of individual creative potential and domain knowledge. *Thinking Skills and Creativity*, *37*, 100682. https://doi.org/10.1016/j.tsc.2020.100682

Suskie, L. (2010). What is good assessment? In *Assessing student learning: a common sense guide* (2nd ed., pp. 36-54). John Wiley & Sons.

Suskie, L. (2018). *Assessing student learning: a common sense guide*. John Wiley & Sons.

Sutton, J., & Austin, Z. (2015). Qualitative research: data collection, analysis, and management. *Can J Hosp Pharm*, *68*(3), 226-231. https://doi.org/10.4212/cjhp.v68i3.1456

Taylor, C. S. (2013). *Validity and validation*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199791040.001.0001

Treffinger, D. J., & Isaksen, S. G. (2005). Creative problem solving: the history, development, and implications for gifted education and talent development. *Gifted Child Quarterly*, 49(4), 342-353. https://doi.org/10.1177/001698620504900407

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., . . . Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: assessment, certification, and promotion of

21st century skills for the future of work and education. *Journal of Intelligence*, *11*(3), 54. https://www.mdpi.com/2079-3200/11/3/54

Tognolini, J. (2018). Assessing skills and measuring performance of students on skills that are aligned with the needs of the 21st Century: a practical example. In *The 2nd University of Riau International Conference on Educational Sciences.*

Tognolini, J., & Davidson, M. (2012). Assessment, standards-referencing and standard setting. In *Self-directed learning oriented assessments in the Asia-Pacific* (pp. 23-41). Springer.

Tognolini, J., & Stanley, G. (2011). A standards perspective on the relationship between formative and summative assessment. *East Asia: Brithis Council*.

Torrance, E. P. (1966a). Nurture of creative talents. *Theory Into Practice*, *5*(4), 168-202. http://www.jstor.org/stable/1475128

Torrance, E. P. (1966b). Torrance test of creative thinking: norms-technical manual. *Personnel Press*.

Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly*, *12*(4), 195-199. https://doi.org/10.1177/001698626801200401

Torrance, E. P. (1969). Predcition of adult creative achievement among high school seniors. *Gifted Child Quarterly*, *13*(4), 223-229. https://doi.org/https://doi.org/10.1177/001698626901300401

Torrance, E. P. (1977). *Creativity in the classroom; what research says to the teacher*. National Education Association.

Torrance, E. P. (1988). The nature of creativity as manifest in its testing. In *The nature of creativity: contemporary psychological perspectives* (pp. 43-75). Cambridge University Press.

Torrance, E. P., & Ball, O. E. (1984). *Torrance tests of creative thinking* (streamlined revised manual ed.). Scholastic Testing Service Inc. Bensenville, Ill.

Towne, L., & Shavelson, R. J. (2002). *Scientific research in education*. National Academy Press Publications.

Tran, L. T. B., Ho, N. T., & Hurle, R. J. (2016). Teaching for creativity development:

lessons learned from a preliminary study of Vietnamese and international upper (high) secondary school teachers' perceptions and lesson plans. *Creative Education*, *7*(07), 1024. https://doi.org/10.4236/ce.2016.77107

Treffinger, D., Isaksen, S., & Stead-Dorval, K. B. (2003). *Creative problem solving (CPS version 6.1™): a contemporary framework for managing change.* Creative Problem Solving Group, Inc.

UNESCO. (2016). Towards 2030: creativity matters for sustainable development. https://unesdoc.unesco.org/ark:/48223/pf0000246746

University of Sydney. (2022). *Graduate qualities*. The University of Sydney. https://www.sydney.edu.au/students/graduate-qualities.html

Urban, K. K. (2004). Assessing creativity: the test for creative thinking - drawing production (TCT-DP): the concept, application, evaluation, and international studies. *Psychology Science*, *46*(3), 387-397.

Urban, K. K., & Jellen, H. (1996). Manual of test for creative thinking-drawing production (TCT-DP). *Amesterdam: Harcourt*.

Vaismoradi, M., Turunen, H., & Bondas, T. (2013). Content analysis and thematic analysis: implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*, *15*(3), 398-405. https://doi.org/https://doi.org/10.1111/nhs.12048

Vincent-Lancrin, S. (2015). *Intervention and research protocol for OECDproject on assessing progression in creative and critical thinking skills in education*. OECD.

Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., de Luca, F., Fernández-Barrerra, M., Jacotin, G., . . . Vidal, Q. (2019). *Fostering students' creativity and critical thinking: what it means in school. Educational research and innovation*. ERIC.

Volle, E. (2018). Associative and controlled cognition in divergent thinking: theoretical, experimental, neuroimaging evidence, and new directions. In R. E. Jung & O. Vartanian (Eds.), *The Cambridge handbook of the neuroscience of creativity* (pp. 333-362). Cambridge University Press.

Vries, H., & Lubart, T. (2017). Scientific creativity: divergent and convergent thinking and the impact of culture. *The Journal of Creative Behavior*, *53*. https://doi.org/10.1002/jocb.184

Vries, H. B., & Lubart, T. I. (2019). *Scientific creativity: divergent and convergent thinking and the impact of culture*. Wiley-Blackwell Publishing Ltd.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children*. New York.

Wallas, G. (1926). *The art of thought*. Harcourt & Brace Company.

Ward, T. B. (2007). Creative cognition as a window on creativity. *Methods*, *42*(1), 28-37. https://doi.org/https://doi.org/10.1016/j.ymeth.2006.12.002

Ward, T. B., Smith, S. M., & Vaid, J. (1997). Conceptual structures and processes in creative thought. In *Creative thought: an investigation of conceptual structures and processes* (pp. 1-27). American Psychological Association. https://doi.org/10.1037/10227-001

Weisberg, R. W. (2006). *Creativity: understanding innovation in problem solving, science, invention, and the arts*. John Wiley & Sons, Inc.

Wilson, H. E., Song, H., Johnson, J., Presley, L., & Olson, K. (2021). Effects of transdisciplinary STEAM lessons on student critical and creative thinking. *The Journal of Educational Research*, *114*(5), 445-457. https://doi.org/10.1080/00220671.2021.1975090

Wilson, M. (2005). *Constructing measures: an item response modeling approach*. Lawrence Erlbaum Associates Publishers.

Wilson, M. (2009). Measuring progressions: assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*(6), 716-730. https://doi.org/https://doi.org/10.1002/tea.20318

Wilson, M. (2018). Making measurement important for education: the crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, *37*(1), 5-20. https://doi.org/https://doi.org/10.1111/emip.12188

Wilson, M., & Lehrer, R. (2021). Improving learning: using a learning progression to coordinate instruction and assessment. *Frontiers in Education.*

Wilson, M., & Scalise, K. (2006). Assessment to improve learning in higher

education: the BEAR assessment system. *Higher Education*, *52*, 635-663. https://doi.org/10.1007/s10734-004-7263-y

Wilson, M., & Sloane, K. (2000). From principles to practice: an embedded assessment system. *Applied Measurement in Education*, *13*(2), 181-208. https://doi.org/10.1207/S15324818AME1302_4

Woolfolk Hoy, A., Davis, H., & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (2nd ed., pp. 715-737). LEA.

Wu, T.-T., & Wu, Y.-T. (2020). Applying project-based learning and SCAMPER teaching strategies in engineering education to explore the influence of creativity on cognition, personal motivation, and personality traits. *Thinking Skills and Creativity*, *35*, 100631. https://doi.org/https://doi.org/10.1016/j.tsc.2020.100631

Xin, X. (2019). *Evaluation of students' higher-order thinking abilities in chemistry in senior high school* [Master's thesis, East China Normal University]. East China Normal University.

Xu, W. (2018). *Research on assessing innovative capacity in high school physics* [Master's thesis, South China Normal University]. South China Normal University.

Yao, X., & Li, R. (2021). Big five personality traits as predictors of employee creativity in probation and formal employment periods. *Personality and Individual Differences*, *182*, 109914. https://doi.org/https://doi.org/10.1016/j.paid.2020.109914

Zhang, Y. (2004). *A foundational study of high school students' creativity in physics* [Doctoral dissertation, Capital Normal University]. Capital Normal University.

Appendices

Appendix A: Project Approval from the University of Sydney Human Research Ethics Committee

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Monday, 18 October 2021

Prof James Tognolini
Social Work; Faculty of Arts and Social Sciences
Email: james.tognolini@sydney.edu.au

Dear James,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that your project has been approved

Details of the approval are as follows:

Project No.:	**2021/774**
Project Title:	**Building and Validating an Assessment Rubric for Measuring Creativity in High School**
Authorised Personnel:	**Tognolini James; Bartimote Kathryn; Xu Weiping;**
Approval Period:	**18/10/2021 to 18/10/2025**
First Annual Report Due:	**18/10/2022**

Documents Approved:

Date Uploaded	Version Number	Document Name
17/09/2021	Version 1	Definition and draft rubric of creativity
17/09/2021	Version 1	Discussion questions for FG1
17/09/2021	Version 1	Discussion questions for FG2
17/09/2021	Version 1	Discussion questions for FG3
17/09/2021	Version 1	Invitation email for FG1
17/09/2021	Version 1	Invitation email for FG2
17/09/2021	Version 2	Invitation for FG3 and teacher workshops (clean)
17/09/2021	Version 2	PCF1_FG 1 and 2 (clean)
17/09/2021	Version 2	PCF2_FG3 and teacher workshops (clean)
17/09/2021	Version 2	PIS1_FG1 (clean)
17/09/2021	Version 2	PIS2_FG2 (clean)
17/09/2021	Version 2	PIS3_FG3 and teacher workshops (clean)

Special Condition/s of Approval

Please correct minor typos in documents before use:

- Phases 3 & 4 of the study will be conducted in January and February 2022, therefore the selection criteria for Group 3 should be updated to "Must be teaching high school physics in 2021 and 2022".
- PIS1FG1: 'standard' should be 'standards' in '(2) Must have a sound knowledge of standards-based assessment'
- PIS2FG2: Change 'If you withdrawal occur after data has been collected,' to 'If your withdrawal occurs ...'
- Invitation FG3: Suggest changing wording of: 'To be eligible you need to have at least two-year teaching experience in schools and being teaching physics in 2021.' to mirror the wording used in PISFG3 i.e. "To be eligible you need to have at least two years' experience of teaching high school physics and be teaching high school physics in 2021'
- 'two online workshop' should be plural ie. 'two online workshops'

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

THE UNIVERSITY OF SYDNEY

- Remove 'working' from the sentence: 'Subsequently, you will work working with other teachers'

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.
- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.
- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or clarification.

Sincerely,

[redacted]

Associate Professor Helen Mitchell
Chair
Human Research Ethics Committee (HREC 1)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2018) and the NHMRC's Australian Code for the Responsible Conduct of Research (2018)

Appendix B: Participant Information Statement (PIS)

1. PIS for Focus Group 1

Participant Information Statement

[Participant group – Focus Group1]

THE UNIVERSITY OF SYDNEY

Building and Validating an Assessment Rubric for Measuring Creativity in High School

Professor Jim Tognolini (Responsible Researcher)
Director, Centre for Educational Measurement and Assessment
Faculty of Arts and Social Sciences
The University of Sydney
Phone: +61 2 8627 5205| Email: jim.tognolini@sydney.edu.au
Miss Weiping Xu (PhD student) | Email: wexu9154@uni.sydney.edu.au

1. What is this study about?

We are conducting a research study about developing and validating an assessment rubric of creativity that when used by teachers in a classroom, will provide the basis for measuring the growth of students' creativity. This rubric will contribute to provide meaningful information on what it is the students know and can do in relation to creativity, and support teachers in identifying creative students in their class.

You have been invited to participate in this study because you have teaching or researching experience in high school. This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the research.

Taking part in this study is voluntary.

Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

2. Who is running the study?

The study is being carried out by the following researchers:

Weiping Xu is conducting this study as the basis for the degree of Doctor of Philosophy at the University of Sydney. This study will take place under the supervision of Professor Jim Tognolini (Chief Investigator) and Dr. Kathryn Bartimote from the Sydney School of Education and Social Work, the University of Sydney.

3. Who can take part in the study?

We are seeking volunteers from amongst academic staff and/or postgraduate research students of the Sydney School of Education and Social Work (SSESW) to participate in the

Focus Group1 (FG1), who have experience teaching or researching in high school contexts. The following are the selection criteria of participants in FG1:
(1) Must have at least one year of experience of teaching or researching in high school classrooms
(2) Must have a sound knowledge of standards-based assessment
(3) Priority will be given to volunteers who have experience in the Chinese context.

You have been invited to take part in this study because your review and comments can benefit the improvement of the generic rubric of creativity, which is part of the validation process. If you are interested and would like to volunteer in this study, you can contact the researcher using the contact information provided. Your contact details will be collected for further communication.

4. What will the study involve for me?

If you decide to take part in this study, you will attend a face-to-face focus group (FG1), which will take approximately 2 hours. You will be asked to review the definition of creativity and its components, as well as the descriptions in the assessment rubric drawing on your experience of teaching and learning in high schools. Your suggestions and comments will be used to revise the generic rubric of creativity assessment.

With your permission, the focus group will be audio-recorded to ensure an accurate record of your comments and suggestions during the discussion. When the recording has been transcribed, you can request a copy of the transcript to be provided, so that you can verify that the information is correct and/or request deletions.

The times for focus group discussion will be negotiated with you to fit with your schedule.

If Covid restrictions allow, FG1 will be a face-to-face discussion. However, if restrictions are still in place at the time, FG1 will be converted to an online discussion.

5. Can I withdraw once I've started?

Being in this study is completely voluntary and you do not have to take part.

Your decision will not affect your current or future relationship with the researchers or anyone else at The University of Sydney.

If you decide to take part in the study and then change your mind you can withdraw by informing any of the researchers via email or telephone using the contact information provided. If you decide to withdraw, we will not collect any more information from you.

If you take part in a focus group, you are free to stop at any stage or to refuse to answer any of the questions. However, since it is a group discussion, it may not be possible to withdraw your individual comments from our records once the group has started.

6. Are there any risks or costs?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

7. Are there any benefits?

Refreshments will be provided during the focus group discussion if held face-to-face.

8. What will happen to information that is collected?

By providing your consent, you are agreeing to us collecting information about you for the purposes of this study.

Any information you provide us will be stored securely and we will only disclose it with your permission, unless we are required by law to release information. We are planning for the study findings to be published.

You will not be individually identifiable in these publications. Pseudonyms or coding by omitting your name will be given to the focus group participants in publications.

By participating the focus group discussions, all your responses to the questions will be audio and some parts of the recordings that is applicable to research objectives will be transcribed. This information will be collected and used for analysis purposes, and where needed, presentations and publications.

Audio files and transcribed data will be stored, saved, and managed using the University of Sydney's Research Data Store(RDS) recommended by the University of Sydney. Secure backup and file storage will adhere to the University of Sydney's Research Data Management Policy. Only the researchers listed in Section 2 above will have the password-protected access to stored data in the RDS.

The study materials will be retained for 5 years after project completion. After 5 years, the hardcopies of the project materials will be disposed and shredded, while the softcopies will be permanently deleted from the storage system.

9. Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the relevant box on the consent form. This feedback will be in the form of a brief lay summary.

10. What if I would like further information?

When you have read this information, the following researcher will be available to discuss it with you further and answer any questions you may have:

Weiping Xu, PhD candidate, email: wexu9154@uni.sydney.edu.au.

11. What if I have a complaint or any concerns?

The ethical aspects of this study have been approved by the Human Research Ethics Committee (HREC) of The University of Sydney [HREC Approval No. 2021/774] according to the *National Statement on Ethical Conduct in Human Research (2007).*

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the University:

Human Ethics Manager
human.ethics@sydney.edu.au
+61 2 8627 8176

This information sheet is for you to keep

2. PIS for Focus Group 2

Participant Information Statement

[Participant group – Focus Group2]

THE UNIVERSITY OF SYDNEY

Building and Validating an Assessment Rubric for Measuring creativity in High School

Professor Jim Tognolini (Responsible Researcher)
Director, Centre for Educational Measurement and Assessment
Faculty of Arts and Social Sciences
The University of Sydney
Phone: +61 2 8627 5205| Email: jim.tognolini@sydney.edu.au
Miss Weiping Xu (PhD student) | Email: wexu9154@uni.sydney.edu.au

1. What is this study about?

We are conducting a research study about developing and validating an assessment rubric of creativity that when used by teachers in a classroom, will provide the basis for measuring the growth of students' creativity. This rubric will contribute to provide meaningful information on what it is the students know and can do in relation to creativity, and support teachers in identifying creative students in their class.

You have been invited to participate in this study because you have experience researching or teaching in science domains within a high school context. This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the research.

Taking part in this study is voluntary.

Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

2. Who is running the study?

The study is being carried out by the following researchers:

Weiping Xu is conducting this study as the basis for the degree of Doctor of Philosophy at the University of Sydney. This study will take place under the supervision of Professor Jim Tognolini (Chief Investigator) and Dr. Kathryn Bartimote from the Sydney School of Education and Social Work, the University of Sydney.

3. Who can take part in the study?

We are seeking volunteers from amongst academic staff and/or postgraduate research students in universities, as well as current school teachers to participate in Focus Group2 (FG2), who have experience teaching or researching in high school science contexts. The following are the selection criteria of participants in FG2:
(1) Must have at least one year experience of teaching or researching in high school science classrooms
(2) Preference will be given to volunteers who have experience in the Chinese context

You have been invited to take part in this study because your review and comments can benefit the improvement of the generic rubric of creativity, which is part of the validation process. If you are interested and would like to volunteer in this study, you can contact the researcher using the contact information provided. Your contact details will be collected for further communication.

4. What will the study involve for me?

If you decide to take part in this study, you will attend a face-to-face focus group (FG2), which will take approximately 2 hours. You will be asked to review the definition of creativity and its components, as well as the descriptions in the assessment rubric through a science-discipline lens. You will comment on whether or not it is necessary to build a domain-specific rubric to enable the measurement of student creativity in the science classroom. Your suggestions and comments will be used to finalise the generic rubric of creativity measurement.

With your permission, the focus group will be audio-recorded to ensure an accurate record of your comments and suggestions during the discussion. When the recording has been transcribed, you can request a copy of the transcript to be provided, so that you can verify that the information is correct and/or request deletions.

The times for the focus group discussion will be negotiated with you to fit with your schedule.

If Covid restrictions allow, FG2 will be a face-to-face discussion. However, if restrictions are still in place at the time, FG2 will be converted to an online discussion.

5. Can I withdraw once I've started?

Being in this study is completely voluntary and you do not have to take part.

Your decision will not affect your current or future relationship with the researchers or anyone else at The University of Sydney.

If you decide to take part in the study and then change your mind you can withdraw by informing any of the researchers via email or telephone using the contact information provided. If you withdrawal occurs after data has been collected, the data (where it can be identified) will be excluded from the study and will remain private and confidential

with all the other data until it is permanently deleted after 5 years upon the completion of the study.

If you take part in a focus group, you are free to stop at any stage or to refuse to answer any of the questions. However, since it is a group discussion, it may not be possible to withdraw your individual comments from our records once the group has started.

If you decide to withdraw, we will not collect any more information from you.

6. Are there any risks or costs?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

7. Are there any benefits?

Refreshment will be provided during the focus group discussion if held face-to-face.

8. What will happen to information that is collected?

By providing your consent, you are agreeing to us collecting information about you for the purposes of this study.

Any information you provide us will be stored securely and we will only disclose it with your permission, unless we are required by law to release information. We are planning for the study findings to be published.

You will not be individually identifiable in these publications. Pseudonyms or coding by omitting your name will be given to the focus group participants in publications.

By participating the focus group discussions, all your responses to the questions will be audio and some parts of the recordings that is applicable to research objectives will be transcribed. This information will be collected and used for analysis purposes, and where needed, presentations and publications.

Audio files and transcribed data will be stored, saved, and managed using the University of Sydney's Research Data Store(RDS) recommended by the University of Sydney. Secure backup and file storage will adhere to the University of Sydney's Research Data Management Policy. Only the researchers listed in Section 2 above will have the password-protected access to stored data in the RDS.

The study materials will be retained for 5 years after project completion. After 5 years, the hardcopies of the project materials will be disposed and shredded, while the softcopies will be permanently deleted from the storage system.

9. Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the relevant box on the consent form. This feedback will be in the form of a brief lay summary.

10. What if I would like further information?

When you have read this information, the following researcher will be available to discuss it with you further and answer any questions you may have:

Weiping Xu, PhD candidate, email:wexu9154@uni.sydney.edu.au.

11. What if I have a complaint or any concerns?

The ethical aspects of this study have been approved by the Human Research Ethics Committee (HREC) of The University of Sydney [HREC Approval No. 2021/774] according to the *National Statement on Ethical Conduct in Human Research (2007).*

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the University:

Human Ethics Manager
human.ethics@sydney.edu.au
+61 2 8627 8176

This information sheet is for you to keep

3. PIS for Focus Group 3 and Teacher Development Workshops

Participant Information Statement

[Participant group – Phase 3 and 4]

THE UNIVERSITY OF SYDNEY

Building and Validating an Assessment Rubric for Measuring Creativity in High School

Professor Jim Tognolini (Responsible Researcher)
Director, Centre for Educational Measurement and Assessment
Faculty of Arts and Social Sciences
The University of Sydney
Phone: +61 2 8627 5205| Email: jim.tognolini@sydney.edu.au
Miss Weiping Xu (PhD student) | Email: wexu9154@uni.sydney.edu.au

1. What is this study about?

We are conducting a research study about developing and validating an assessment rubric of creativity that when used by teachers in a classroom, will provide the basis for measuring the growth of students' creativity and the impact of their teaching and student learning on the development of it. This rubric will contribute to provide meaningful information on what it is the students know and can do in relation to creativity, and support teachers in identifying creative students in their class.

You have been invited to participate in this study because you are an experienced physics teacher. This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the research.

Taking part in this study is voluntary.

Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

2. Who is running the study?

The study is being carried out by the following researchers:

Weiping Xu is conducting this study as the basis for the degree of Doctor of Philosophy at the University of Sydney. This study will take place under the supervision of Professor Jim Tognolini (Chief Investigator) and Dr. Kathryn Bartimote from the Sydney School of Education and Social Work, the University of Sydney.

3. Who can take part in the study?

We are seeking volunteers from amongst physics teachers in Chinese high schools to participate in the Focus Group3 (FG3) as well as a series of two teacher development workshops. The following are the selection criteria of participants in FG3:
(1) Must have at least one years experience of teaching high school physics
(2) Must be teaching high school physics in 2021 and 2022
(3) Be located in Guangzhou, Shenzhen and Foshan districts

You have been invited to take part in this study because your review and comments can benefit the improvement of the generic rubric of creativity, which is part of the validation process. If you are interested and would like to volunteer in this study, you can contact the researcher using the contact information provided. Your contact details will be collected for further communication.

4. What will the study involve for me?

If you decide to take part in this study, you will attend an online focus group (FG3) and two online teacher development workshops.

The FG3 will take approximately 2 hours, where you will be asked to review the definition of creativity and its components, as well as the descriptions in the assessment rubric through a physics-discipline lens. You will comment on whether or not it is necessary to build a domain-specific rubric in physics to enable the measurement of student creativity in the physics classroom. Your suggestions and comments will be used to refine and determine the assessment rubric of creativity that will be applied in the following teacher development workshops.

There are two online teacher development workshops, which will be held over a period of 2 weeks. They aim to demonstrate the applicability of the defined rubric in classroom assessment. The first workshop will be for approximately 2 hours, which is designed to introduce background knowledge on creativity assessment as well as approaches to design tasks against the rubric. The second workshop will last for around 3 hours, where you will be asked to design some tasks against the rubric that can be used to measure creativity in physics teaching, and to peer review these tasks.

With your permission, the focus group discussion and the workshops will be audio-recorded to ensure an accurate record of your comments and suggestions during the discussion. When the recording has been transcribed, you can request a copy of the transcript to be provided, so that you can verify that the information is correct and/or request deletions.

The times for the focus group discussion and the workshops will be negotiated with you to fit with your schedule.

5. Can I withdraw once I've started?

Being in this study is completely voluntary and you do not have to take part.

Your decision will not affect your current or future relationship with the researchers or anyone else at The University of Sydney.

If you decide to take part in the study and then change your mind you can withdraw by informing any of the researchers via email using the contact information provided. If you decide to withdraw, we will not collect any more information from you.

If you take part in a focus group or a series of teacher development workshops, you are free to stop at any stage or to refuse to answer any of the questions. However, since it is a group discussion, it may not be possible to withdraw your individual comments from our records once the group has started.

6. Are there any risks or costs?

Aside from giving up your time, we do not expect that there will be any risks or costs associated with taking part in this study.

7. Are there any benefits?

You will not receive any direct benefits from being in the study.

8. What will happen to information that is collected?

By providing your consent, you are agreeing to us collecting information about you for the purposes of this study.

Any information you provide us will be stored securely and we will only disclose it with your permission, unless we are required by law to release information. We are planning for the study findings to be published.

You will not be individually identifiable in these publications. Pseudonyms or coding by omitting your name will be given to the focus group participants in publications.

By participating the focus group discussions and teacher development workshops, all your responses to the questions will be audio and some parts of the recordings that is applicable to research objectives will be transcribed. This information will be collected and used for analysis purposes, and where needed, presentations and publications.

Audio files and transcribed data will be stored, saved, and managed using the University of Sydney's Research Data Store(RDS) recommended by the University of Sydney. Secure backup and file storage will adhere to the University of Sydney's Research Data Management Policy. Only the researchers listed in the Section 2 above will have the password-protected access to stored data in the RDS.

The study materials will be retained for 5 years after project completion. After 5 years, the hardcopies of the project materials will be disposed and shredded, while the softcopies will be permanently deleted from the storage system.

9. Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by ticking the relevant box on the consent form. This feedback will be in the form of a brief lay summary.

10. What if I would like further information?

When you have read this information, the following researcher will be available to discuss it with you further and answer any questions you may have:

Weiping Xu, PhD candidate, email:wexu9154@uni.sydney.edu.au.

11. What if I have a complaint or any concerns?

The ethical aspects of this study have been approved by the Human Research Ethics Committee (HREC) of The University of Sydney [HREC Approval No. 2021/774] according to the *National Statement on Ethical Conduct in Human Research (2007).*

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the University:

Human Ethics Manager
human.ethics@sydney.edu.au
+61 2 8627 8176

This information sheet is for you to keep

Appendix C: Participant Consent Form (PCF)

1. PCF for FG1 and FG2

Participant Consent Form

Focus Group 1 and Focus Group 2

THE UNIVERSITY OF SYDNEY

Research Study: Building and Validating an Assessment Rubric for Measuring Creativity in High School

Professor Jim Tognolini (Responsible Researcher)
Director, Centre for Educational Measurement and Assessment
Faculty of Arts and Social Sciences
The University of Sydney
Phone: +61 2 8627 5205| Email: jim.tognolini@sydney.edu.au
Miss Weiping Xu (PhD student) | Email: wexu9154@uni.sydney.edu.au

---

Participant Name _____

I agree to take part in this research study. In giving my consent, I confirm that that:

- The details of my involvement have been explained to me, and I have been provided with a written Participant Information Statement to keep.
- I understand the purpose of the study is to develop and validate an assessment rubric of creativity that can provide meaningful information on what it is the students know and can do in relation to creativity.
- I acknowledge that the risks and benefits of participating in this study have been explained to me to my satisfaction.
- I understand that in this study I will be required to discuss the focus group questions with my peers in a face-to-face focus group (if Covid restrictions allow).
- I understand that my participation will be audio-taped.
- I understand that being in this study is completely voluntary.
- I am assured that my decision to participate will not have any impact on my relationship with the research team or the University of Sydney.
- I understand that I am free to withdraw from this study at any time and that I can choose to withdraw any information I have already provided (unless the data has already been de-identified or published).
- I have been informed that the confidentiality of the information I provide will be protected and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.

- I understand that the results of this study may be published, and that publications will not contain my name or any identifiable information about me.
- I confirm the following:

I consent to audio-recordings	Yes ☐	No ☐
I would like to review my meeting transcripts	Yes ☐	No ☐
I would like feedback on the overall results of this study	Yes ☐	No ☐

If you answered **yes**, please provide your preferred contact details (email/telephone/postal address):

_____

_____

_____

_____

- I understand that after I sign and return this consent form it will be retained by the researcher, and that I may request a copy at any time.

Participant Name _____

Signature _____

Date _____

2. PCF for FG3 and Teacher Development Workshops

Participant Consent Form

Focus Group 3 and teacher development workshops

THE UNIVERSITY OF SYDNEY

Research Study: Building and Validating an Assessment Rubric for measuring creativity in High School

Professor Jim Tognolini (Responsible Researcher)
Director, Centre for Educational Measurement and Assessment
Faculty of Arts and Social Sciences
The University of Sydney
Phone: +61 2 8627 5205| Email: jim.tognolini@sydney.edu.au
Miss Weiping Xu (PhD student) | Email: wexu9154@uni.sydney.edu.au

Participant Name _____

I agree to take part in this research study. In giving my consent, I confirm that that:

- The details of my involvement have been explained to me, and I have been provided with a written Participant Information Statement to keep.
- I understand the purpose of the study is to develop and validate an assessment rubric of creativity that can provide meaningful information on what it is the students know and can do in relation to creativity.
- I acknowledge that the risks and benefits of participating in this study have been explained to me to my satisfaction.
- I understand that in this study I will be required to discuss the focus group questions with my peers in ~~an online~~ a face-to-face focus group discussion and to validate the applicability of the measurement rubric in ~~a series of~~ two offline teacher development workshops (if Covid restrictions allow).
- I understand that my participation will be audio-taped.
- I understand that being in this study is completely voluntary.
- I am assured that my decision to participate will not have any impact on my relationship with the research team or the University of Sydney.
- I understand that I am free to withdraw from this study at any time and that I can choose to withdraw any information I have already provided (unless the data has already been de-identified or published).
- I have been informed that the confidentiality of the information I provide will be protected and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.

- I understand that the results of this study may be published, and that publications will not contain my name or any identifiable information about me.
- I confirm the following:

I consent to audio-recordings	Yes ☐	No ☐
I would like to review my meeting transcripts	Yes ☐	No ☐
I would like feedback on the overall results of this study	Yes ☐	No ☐

If you answered **yes**, please provide your preferred contact details (email/telephone/postal address):

_____

_____

_____

_____

- I understand that after I sign and return this consent form it will be retained by the researcher, and that I may request a copy at any time.

Participant Name _____

Signature _____

Date _____

Appendix D: FG1 Discussion Questions

Validation Phase 1 – Validate the generic measurement rubric for creativity across different disciplines

Questions:

1. What do you think about the definition of creativity? Is the definition appropriate?
2. What do you think about the components provided in the drafted rubric? Are the components aligned with the definition? Do any components need to be reworded or eliminated because they are irrelevant to the definition of creativity? Are there other components necessary to give a more faithful description of the definition?
3. To what extent are the performance statements in the rubric appropriate and explicit enough to clarify the developmental performance of student creativity at different levels? To what extent are the statements easy for you to understand? Does the rubric give you an image of different levels of creativity that could help you to recognise student creativity in your classroom?
4. Can you describe some examples of students displaying either creative potential or creative achievement against the components at different levels?
5. (If applicable), can you please describe any tasks you have used to assess student creativity?

Appendix E: FG2 Discussion Questions

Validation Phase 2 – Validate the generic measurement rubric for creativity in the domain-specific context of science.

Questions:

1. What do you think about the definition of creativity? Is the definition appropriate?
2. What do you think about the components provided in the drafted rubric? Are the components aligned with the definition? Do any components need to be reworded or eliminated because they are irrelevant to the definition of creativity? Are there other components necessary to give a more faithful description of the definition?
3. Are the performance statements in the rubric appropriate and explicit enough to clarify the developmental performance of student creativity at different levels? Are the statements easy for you to understand? Does the rubric give you an image of different levels of creativity that could help you to recognise student creativity in your classroom?
4. Can you describe some examples of students displaying either creative potential or creative achievement against the components at different levels?
5. (If applicable), can you please describe any tasks you have used to assess student creativity?
6. Is it necessary to build a domain-specific rubric to enable the measurement of student creativity in the science classroom? If it is necessary, what changes would you make to the words in the generic rubric to make it a better fit in the science context?

Appendix F: FG3 Discussion Questions

Validation Phase 3 – Validate the generic measurement rubric for creativity in the domain-specific context of physics in China.

Questions:

1. Are the performance statements in the rubric appropriate and explicit enough to clarify the developmental performance of student creativity at different levels? Are the statements easy for you to understand? Does the rubric give you an image of different levels of creativity that could help you to recognise student creativity in your classroom?
2. Is it necessary to build a domain-specific rubric that to make it easier to evaluate student creativity in the physics classroom? If it is necessary, what changes would you make to the words in the generic rubric to make it a better fit in the physics context?
3. Can you describe some examples of students displaying either creative potential or creative achievement against the components at different levels?
4. (If applicable), can you please describe any tasks you have used to assess student creativity?

Appendix G: Agenda of Two Teacher Development Workshops

1. Agenda for Teacher Development Workshop 1

Duration	Topic	Content
15 mins	Project overview	Introducing the rationale of the research project and the goal of two teacher development workshops
30 mins	Theories for a measurement rubric	1) Modern measurement theory: measuring growth of a construct 2) Standards-referenced model: referencing student response to a set of predetermined performance-level descriptors 3) Compositions of a measurement rubric and their meaning in assessment
15 mins	Application of a measurement rubric in science	1) Using Australian Curriculum in Science as example: illustrating a clear alignment between items, the performance of the task elicits from the student and the performance-level descriptors 2) Teachers' attempts to design tasks based on the performance standards in Australian Science Curriculum
		Break (15 mins)
20 mins	Task principles for measuring HOT	Discussing the implications of each principle by providing examples of tasks
25 min	Group discussion	Validating existing items measuring creativity against the measurement rubric and principles

2. Agenda for Teacher Development Workshop 2

Duration	Topic	Content
10 mins	Project overview	Overviewing the task design principles and the generic measurement rubric for creativity
20 mins	Drafting tasks	Teachers drafting their task based on existing items of measuring creativity and their teaching experience
20 mins	1st peer review	Teachers sharing their tasks with each other and providing comments by referencing to the generic measurement rubric for creativity and the principles
Break (10 mins)		
15 mins	Revising tasks	Teachers revising their tasks according to the peer-review feedback
20 mins	2nd peer review	Teachers reviewing the revised tasks and examining whether the issues provided during the 1st peer review have been resolved
10 mins	Finalising tasks	Teachers refining and finalising their tasks according to the peer-review feedback
15 mins	Group discussion	Was there a need for a task-specific rubric?