Thinking Skills and Creativity 17 (2015) 1–14

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: http://www.elsevier.com/locate/tsc

Building a learning progression for scientific imagination: A measurement approach

Chia-Chi Wang, Hsiao-Chi Ho, Ying-Yao Cheng*

National Sun Yat-Sen University, Taiwan

ARTICLE INFO

Article history:
Received 4 October 2014
Received in revised form 29 January 2015
Accepted 18 February 2015
Available online 26 March 2015

Keywords:
Scientific imagination
Learning progression
BEAR Assessment System
Rasch model
Elementary school level

ABSTRACT

This study aimed to build a learning progression (LP) for the development of scientific imagination based on a measurement approach using the BEAR Assessment System (BAS) in an attempt to better understand the core ideas and the developmental path of the scientific imagination process as well as align curriculum, instruction, and assessment through LP. Participants in this study were selected from Taiwan, and classified into two categories. The first category included 741 5th and 6th grade elementary school students, were administered the Scientific Imagination Test-Verbal (SIT-Verbal). The second category included one award-winning teacher who designed and implemented a set of curriculum for scientific imagination. The SIT-Verbal was developed by a panel of experts and covered four key components of scientific imagination process: brainstorming, association, transformation/elaboration, and conceptualization/organization/formation. The multiple validities of the SIT-Verbal were assessed using the Rasch partial credit model. Results showed that the components of brainstorming, association, and transformation/elaboration were hierarchical. Additionally, the SIT-Verbal was suitable for measuring students' scientific imagination at the elementary school level. Based on the proposed LP, a set of science curricula was developed for science classroom. The teacher's reflections and observations of LP application were recorded to provide insight into scientific imagination development in practice. The conclusion of the study not only enhance teachers' professions, but also provide more abundant information to verify the LPs for scientific imagination. Implications for the assessments with the LPs and revisions for both the SIT-Verbal and the scientific imagination LP are also proposed.

1. Introduction

Imagination and innovation are key elements driving the present economy and culture (McCormack, 2010). The development of technology today depends on implementing the best strategies for turning imagination into creativity (Vygotsky, 1930/2004). Many scientific theories and inventions have originated primarily from ideational processes within what is commonly referred to as the human imagination. For example, the 19th-century German chemist August Kekulé pictured the ring structure of benzene after dreaming of a snake eating its own tail; this discovery provided a solution to several difficult problems at the time, thus implying that dream images could translate into chemical reality (Robinson, 2010). In the classic

* Corresponding author at: Institute of Education, National Sun Yat-sen University, 70 Lienhai Rd., Kaohsiung 80424, Taiwan. Tel.: +886 7 5252000x5856; fax: +886 7 5255892.
E-mail address: chengyy@mail.nsysu.edu.tw (Y.-Y. Cheng).

http://dx.doi.org/10.1016/j.tsc.2015.02.001

Chinese novel *Journey to the West* set in the Ming dynasty, the Monkey King, Sun Wu-Kong, who created clones of himself from his own hair was echoed in the cloning of Dolly the sheep in the 20th century (Campbell, McWhir, Ritchie, & Wilmut, 1996). Another example is Albert Einstein, who having famously imagined himself flying at light speed and visualizing the objects that he might see, ultimately developed the theory of general relativity. The "invisibility cloak" in the Harry Potter series may become a reality, as metamaterials are currently being investigated by scientists (Pendry, Schurig, & Smith, 2006). As depicted in the examples provided, we 'see' through operational processes, manipulations, and interactions involving the imagination; a human being recognizes internally generated creative ideas that lead ultimately to the invention or design of concrete objects that are eventually manufactured into products (Eckhoff & Urbach, 2008).

Imagination has a great influence on people's thinking, language, and life experiences (Adams, 2004; Grant, 2004; Mountain, 2007). Processes stemming from human imagination potentially provide people with opportunities to explore the world, follow their interests, find answers to problems, and further develop capabilities that are necessary for future survival (Adams, 2004; Church, 2006; Grant, 2004; Mey, 2006; Mountain, 2007; Osborn, 1953; Vygotsky, 1930/2004; Zabriskie, 2004; Zhao, Hoeffler, & Dahl, 2009). Today's science education is the best opportunity for emphasizing imagination and innovation (McCormack, 2010). Infusing imagination into science education (e.g., student-centered scientific studies or innovation venues) would benefit the entire curriculum by deepening and broadening students' scientific concepts. Only by exerting imagination can we surpass existing knowledge and extend beyond the limitations of experience to produce new ideas for solving problems (Church, 2006).

In April 2013, the National Research Council of the United States created a framework for kindergarten through 12th grade science education, referred to as the Next Generation Science Standards (NGSS). Specifically, these standards are based on three dimensions: (1) disciplinary core ideas, (2) science and engineering practices, and (3) crosscutting concepts. The NGSS reflect an evolved vision of inquiry-based learning, emphasizing science as a knowledge-building endeavor. An improvement over prior science education standards, the NGSS is embedded in LPs research-based cognitive models of how learning of scientific concepts and practices unfolds over time (Duncan & Rivet, 2013). Moreover, in the NGSS report, one category in the appendix Nature of Science, "Science is a Human Endeavor," emphasizes the importance of imagination and creativity in different phases of elementary school grades 3–5, middle school, and high school (The Next Generation Science Standards, 2013). The importance of higher-order thinking skills, such as imagination and creativity, in science education grows and changes over time and requires attention from an early age.

In recent years, LPs have been widely discussed in science education. LPs are successively more sophisticated ways of thinking about a topic or big ideas over an extended period of time and can be used as templates for the development of curriculum and assessments (Smith, Wiser, Anderson, & Krajcik, 2006; Wilson, 2009). LPs have also been described as a conjectural model of learning over time that still requires empirical validation (Duncan, 2009). Recent policy reports and studies have advocated for the use of LPs as a means of aligning curriculum, instruction, and assessment (NRC, 2006, 2007; Smith et al., 2006; Wilson, 2009). Moreover, the Journal of Research in Science Teaching from the National Association for Research in Science Teaching (NARST) published a special issue about "Learning Progressions" in 2009 (Duncan, 2009). Thus, LPs studies have become an important trend among researchers worldwide, requiring extensive exploration from various cultural contexts.

More recently, LPs have been used to explore "the big ideas" in scientific disciplines. Examples include the food chain concept (Gotwals & Songer, 2010; Songer & Gotwals, 2012), biodiversity (Songer, Kelcey, & Gotwals, 2009), and genetics (Duncan & Tseng, 2011; Duncan, Rogat, & Yarden, 2009; Freidenreich, Duncan, & Shea, 2011) in biology; celestial motion in earth science (Plummer & Krajcik, 2010; Wilson, 2009); force and motion in physics (Alonzo & Steedle, 2009; Fulmer, Liang, & Liu, 2014; Steedle & Shavelson, 2009); matter in chemistry (Adadan, Trundle, & Irving, 2010; Liu & Lesniak, 2006; Johnson & Tymms, 2011; Stevens, Delgado, & Krajcik, 2010); and energy (Lee & Liu, 2010) and carbon cycling (Mohan, Chen, & Anderson, 2009) across disciplines. However, in terms of LPs, higher-order thinking is rarely mentioned in non-disciplines, with the exception of "scientific modeling" (Schwarz, Reiser, Davis, & Kenyon, 2009) and "scientific argumentation" (Berland & McNeill, 2010) in science education. In fact, imagination is an important issue in science field; for instance, scientific imagination is the desire to deal with inconveniences encountered in daily life, and problem-solving also depends on the operation of imagination (Ho, Wang, & Cheng, 2013). Therefore, the purpose of this study was to build a LP for scientific imagination based on the perspective of scientific invention in the informal activities of science education and attempt to understand the core ideas and learning paths of the scientific imagination process for primary school students through feedback from the LPs.

2. Building a LP for scientific imagination: alignment of curriculum, instruction, and assessment

2.1. Building a LP for scientific imagination: a measurement approach

In the past, most LPs researches (e.g., Alonzo & Steedle, 2009; Claesgens, Scalise, Wilson, & Stacy, 2009; Duncan & Hmelo-Silver, 2009; Mohan et al., 2009; Stevens et al., 2010) indicated that the development of LPs was an iterative process. In this process, the researchers first propose a hypothetical theoretical model of the LPs by exploring the main concepts in specific disciplines and reviewing the literature. Then, empirical data are collected to verify and modify the hypothetical model iteratively. Common methods for exploring LPs include exploring the different levels of students' understanding through assessment (Johnson & Tymms, 2011; Songer & Gotwals, 2012), clarifying students' level of understanding through interviews

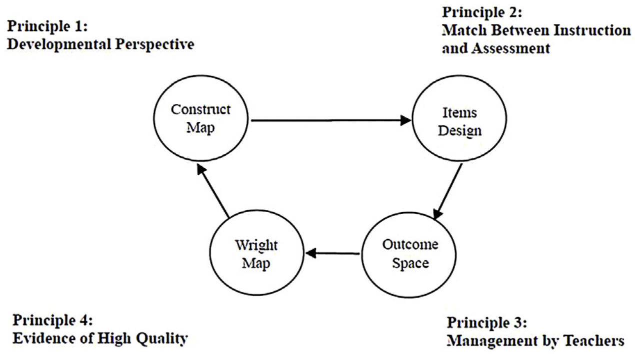

Fig. 1. Four building blocks in the Berkeley Evaluation and Assessment Research (BEAR) assessment system (BAS), from "Measuring progression: Assessment structures underlying a learning progression", by M. Wilson, 2009, Journal of Research in Science Teaching, 45 (6), 718.

(Schwarz et al., 2009; Stevens et al., 2010), comparing the pre-teaching and post-teaching capabilities of students through conceptual application/implementation (Adadan et al., 2010; Duncan & Tseng, 2011; Freidenreich et al., 2011; Plummer & Krajcik, 2010), comparing students' level of understanding at different ages (Lee & Liu, 2010; Mohan et al., 2009; Steedle & Shavelson, 2009), and analyzing changes in students' level of understanding using longitudinal research studies (Songer et al., 2009).

Although these methods provide information helpful for LP instruction, these past studies rarely mention how to implement the LPs in practice. To align LPs development with curriculum and assessment, the Berkeley Evaluation and Assessment Research Center (BEAR) developed the BEAR Assessment System (BAS). The BAS combines the functions of formative and summative assessment to diagnose students' performance and efficiency using feedback (Wilson, 2005). This approach has been used in several recent studies to assess students' conceptual knowledge of matter (Claesgens et al., 2009), celestial motion (Briggs, Alonzo, Schwab, & Wilson, 2006), and force and motion (Alonzo & Steedle, 2009). The creation of construct maps and design items using the BAS facilitates student assessment (Wilson, 2009) and provides useful feedback to enhance teachers' pedagogical strategies (Black, Wilson, & Yao, 2011). Therefore, in the current study, BAS was applied to the development of a LP for scientific imagination that aligns scientific invention curriculum, instruction, and assessment.

The BAS has four principles: (1) a developmental perspective, (2) a match between instruction and assessment, (3) the generating of quality evidence, and (4) management by instructors to allow appropriate feedback, feed forward and follow-up (Wilson, 2009). The BAS transforms these four principles into four concrete building blocks, as shown in Fig. 1: development progress variables, item design, outcome space, and Wright mapping (measurement model) (Wilson, 2005, 2009). The current study applied the BAS to build a LP for scientific imagination (Fig. 1).

2.1.1. Development progress variables

The development progress variables are the variables or constructs corresponding to the understanding level of the student over the duration of the LPs, which are assessed as a means of gauging progress. To clarify students' understanding level in one or more categories, hierarchies of conceptual understanding or *construct maps* (Wilson, 2005, 2009) can be created. Researchers can then use the construct maps to design various types of assessments. Here, the development progress variable would be the construct of scientific imagination.

Researchers have found it quite difficult to clearly define the influential factors and processes leading to imagination and inspiration, as evidenced by the rare mention of imagination processes. In science field, Cruz and Smedt (2010) explored what enables scientists to make significant contributions, meaningful scientific theories, and scientific inventions that influence human civilization. They observed that imagination is a crucial, fundamental factor in the creation process. The tendency to rely on existing knowledge as a guide to creativity is termed *structured imagination* (Cruz & Smedt, 2010; Ward, 1994). Imagination in science field is "structured," as most scientific progress taking place in mental "hops" rather than great leaps (Cruz & Smedt, 2010).

However, it was still not clear to identify the cognitive components of imagination in science field. Recently, Ho et al. (2013) proposed that imagination in scientific fields was based on purposeful imagination, they named it *scientific imagination*. The definition of scientific imagination was a mental activity involving the creation of new ideas that are consistent with scientific principles linked to daily life experiences. In their study, Ho et al. (2013) explored scientific imagination by studying award-winning teachers and students in the International Exhibition for Young Inventors (IEYI) competition. They identified three key stages were in the scientific imagination process: initiation, dynamic adjustment, and virtual implementation. During these three stages of the scientific imagination process, four underlying components were emphasized: brainstorming (*Man-hsiang*, 漫想), association (*Lien-hsiang*, 聯想), transformation/elaboration (*Chi-hsiang*, 奇想), and conceptualization/organization/formation (*Miao-hsiang*, 妙想).

Based on the process of scientific imagination identified by Ho et al. (2013), we invited several experts who specialized in educational testing and assessment, science education, and imagination, as well as experienced teachers, to discuss the

Table 1
LP of scientific imagination as a tool for indicators.

Scientific imagination	Stages	Abilities	Indicators
	3. Virtual Implementation Stage	3.1 Conceptualization/ Organization/Formation	3.1 Students can refine their previous activities to create a prototype for their specific ideas, including issues related to the choice of materials, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams
	2. Dynamic Adjustment Stage	2.3 Transformation/Elaboration	2.3 Student can reorganize the physical features and functions of creations to solve problems
		2.2 Association	2.2 Students can do with possible solutions and illustrate the physical features and functions of their creations
		2.1 Brainstorming	2.1 Based on the problems, student can propose solutions or creations
	1. Initiation Stage	1.2 Association	1.2 The problems students specify have to correlate with life experience
		1.1 Brainstorming	1.1 Based on the question situation, students can specify what problems (e.g., inconvenience) the situation might bring about

construct of scientific imagination over three discussion periods. Although each of the three stages in the process had its main components from the four given above, the experts indicated that some components may also work in other stages in a different way (see Table 1). For example, brainstorming (*Man-hsiang*, 漫想) refers to thinking in such a way as to generate numerous ideas without regard for the usual boundaries or structures. The ideas generated may be in the form of associated problems familiar to the student or solutions that they propose. Brainstorming in the initiation stage focuses on the number of problems students propose; however, in the dynamic adjustment stage, brainstorming involves proposing as many solutions as possible. Second, association (*Lien-hsiang*, 聯想) refers to finding ways to envision relationships among ideas. Students connect related ideas, extend concepts behind ideas, identify contradictions among ideas, and reorganize them accordingly. In the initiation stage, for example, association emphasizes generating as many relationships among ideas as possible. However, in the dynamic adjustment stage, association has to do with possible solutions and expected function/outcome. Transformation/elaboration (*Chi-hsiang*, 奇想) involves transforming an emergent idea into a novel idea by exploring its associative network. Student can reorganize the appearance physical features and functions of creations. Finally, conceptualization/organization/formation (*Miao-hsiang*, 妙想) has to do with how students refine their previous activities to create a prototype for their specific ideas, including issues related to the choice of materials, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams. The resulting prototype can then be used as the basis for scientific imagination and can be linked to subsequent creations; this ability only works in the virtual implementation stage. This process was similar to the cyclic framework suggested by Wilson (2009) that could be carried out on site in the educational setting.

2.1.2. *Item design*

Item design governs the match between classroom instruction and the various types of assessments (Wilson, 2009). Based on the first step, we can design and develop various tools or tests to collect students' actual responses as they study.

The Scientific Imagination Test-Verbal (SIT-Verbal) was designed in this study specifically for 5th and 6th grade elementary school students, is a situation test that measures students' scientific imagination. Five experts specializing in imagination, creativity, science education, and educational testing and assessment were invited to view the item contents according to the construct of scientific imagination. The items were revised according to suggestions of these experts. The content of the test was about a space mission to another planet in which the protagonists (the students) encountered "a lot of leaves at school." In the open-ended test, students were required to complete two missions (Table 2). The first mission (Mission 1) consisted of two questions. The first question required the participants to first identify the potential problems given by the situation "a lot of leaves at school." The two components associated with this question (Question 1) were brainstorming and association. The second question (Question 2) asked students to design solutions to the problem. The components included in the second question were brainstorming, association, and transformation/elaboration. Students were instructed to include descriptions and illustrations for both questions, as well as operating instructions for their solution. Students were given 7 min to complete the first mission.

Following successful completion of the first mission (answering the first two questions), the students were given a second mission (Mission 2), in which they had to draw one "new invention" for solving the problem identified by Question 1 in the first mission (Mission 1). Time was limited to 10 min for this portion of the test. For Mission 2, only the conceptualization/organization/formation component was included. The test covered six items related to the four components.

Table 2
Content included in the Scientific Imagination Test-Verbal (SIT-Verbal).

Mission 1
The planet LABIDO has rich natural resources and minerals which could help to solve crises facing the earth. However, the LABIDO is in a big trouble now. In order to help the LABIDO, you are chosen as one of spacemen to take the adventure. Before your journey, you need to pass one mission. Then you can go to the LABIDO and complete the second mission with other spacemen.
Situation:
"*There is a school on the LABIDO. Many trees are planted in the school. There are always lots of falling leaves near the trees*. . .." Please answer following questions. Limited time: 7 min
1. Question: What problems might the situation bring about? (the more problems you can think about, the better)
2. Question: How many solutions or creations could you think of and illustrate how to do? (the more solutions, the better)
Mission 2
Following the same situation, Please draw one "new invention" that you think can solve the problem in Situation efficiently. You have to explain what materials you will need for your new invention and specify the functions of this invention. Please also give your invention a name. Limited time: 10 min

2.1.3. Outcome space

Outcome space is the set of categorical outcomes into which students' performances are categorized for all items associated with a progress variable (Wilson, 2009). In practice, scoring guides for student responses to assessment tasks are necessary for teachers. From these assessments, teachers can diagnose students' understanding at particular levels to better understand their learning aptitude based on typical responses. The scoring guides for student responses were drawn up through expert consensus from the three meetings/discussions and pretest results. The quality of the students' responses was ranked from "0" (Level 0: the lowest level) to "3" (Level 3: the highest level) (Table 3).

Five experts rated the participants' responses. This group of experts included three outstanding primary school teachers from Taiwan, who had more than 23 combined years of teaching experience in guiding international inventive competitions and science fairs (two males and one female), and two doctoral candidates, who majored in educational psychology, creativity, imagination, and educational testing and assessment, and had published several journal papers related to these fields (e.g., Ho et al., 2013; Wang, Cheng, Liu, & Ho, 2011; Wang, Ho, Cheng, & Cheng, 2014a; Wang, Ho, Wu, & Cheng, 2014b). Once the group of experts reached consensus on the scoring guide, they were each given the same set consisting of 10 questionnaires for rating; this allowed the inter-rater reliability to be determined. The reliability values varied from .41 to .86, indicating stability among raters in rating the 10 questionnaires in the set.

2.1.4. Measurement model

The BAS uses Rasch models to analyze data. Rasch models are commonly used to represent participants' responses in terms of probabilities, which may overcome some limitations of psychological testing associated with classical test theory (CTT). For example, if the observed responses are in agreement with the model, then the ordinal scale can be converted to an interval scale using Rasch analysis. Additionally, because Rasch measurement is based on theoretical foundations, it can avoid inconsistent factor structures arising from exploratory factor analysis. Moreover, Rasch analysis not only examines theoretical constructs but also provides information regarding item hierarchies and latent traits. Overall, the quality of an assessment tool can be tested using multiple sources of evidence offered by the Rasch measurement approach to better understand the constructs (e.g., scientific imagination).

In the current study, because each item of the SIT-Verbal has four response categories, the Rasch partial credit model (Rasch PCM; Masters, 1982) was employed to fit to the data. The PCM is a unidimensional model for the analysis of multiple ordered responses. The reason we used PCM rather than other IRT model such as GPCM is that GPCM is one kind of 2-parameter logistic models. The abilities estimated from 2-parameter logistic models do not have the objective and interval characteristics. Only 1-parameter logistic models (Rasch models) possess these characteristics.

This Rasch PCM consists of two primary parameters: the person's (n) ability (θ_n) and the item's (i) difficulty (δ_i). When person n responds to item i, the probability of that person's having the correct answer to that item is given as follows:

$$P_{nix} = \frac{\exp \sum_{j=0}^{x} [\theta_n - (\delta_i + \tau_{ij})]}{\sum_{k}^{m_i} \exp \sum_{j=0}^{k} [\theta_n - (\delta_i + \tau_{ij})]}, \quad x = 0, 1, \ldots, m_i \tag{1}$$

where P_{nix} is the probability of person n's scoring x on item i. Additionally, θ_n, which is the latent trait level of person n, refers to the construct that is the target of the measurement (e.g., personality, attitudes, interests, values, performance, or in this case, scientific imagination), and δ_j is the overall difficulty of item i (e.g., difficulty or threshold value). τ_{ij} is an additional step parameter of scoring j rather than $j-1$ on item i. To compare the hierarchies of scientific imagination, this study calculated the Thurstone thresholds of each item. Because the SIT-Verbal uses a four-point scale, each item has three Thurstone threshold values. The kth Thurstone threshold of each item of the SIT-Verbal corresponds to the boundary point at which the probability of the participant's getting a score below k is equal to the probability of that participant's getting a

Table 3
Summaries of matched indicators and items' descriptions for scientific imagination.

Stages	Abilities	Scoring guides for student responses	Description of items
3. Virtual Implementation Stage	3.1 Conceptualization/ Organization/Formation	3.1 Students can refine their previous activities to create a prototype for their specific ideas, including issues related to the choice of materials, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams 3.1.1 Students cannot consider the possibilities of ideas and draw them on the paper, including issues related to the choice of materials, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams. (Level 0) 3.1.2 Students can roughly propose the idea of creation through speaking or writing. (Level 1) 3.1.3 Students can propose the idea of creation through speaking or writing and roughly consider the possibilities of ideas and draw them on the paper, including the choice of material, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams. (Level 2) 3.1.4 Students can draw a detailed draft of creation and illustrate details for their choice of material, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams. (Level 3)	Mission 2
2. Dynamic Adjustment Stage	2.3 Transformation/ Elaboration	2.3 Student can reorganize the physical features and functions of creations to solve problems. 2.3.1Students cannot reorganize the physical features and functions of creations to solve problems. (Level 0) 2.3.2 Students can reorganize one physical feature and function of creations to solve problems. (Level 1) 2.3.3 Students can reorganize two physical features and functions of creations to solve problems. (Level 2) 2.3.4 Students can reorganize at least three physical features and functions of creations to solve problems. (Level 3)	Mission 2
	2.2 Association	2.2 Students can do with possible solutions and illustrate the physical features and functions of their creations 2.2.1 Students cannot illustrate and functions of their creations. (Level 0) 2.2.2Students can illustrate one physical feature and function of their creations. (Level 1) 2.2.3Students can illustrate two physical features and functions of their creations. (Level 2) 2.2.4 Students can illustrate at least three physical features and functions of creations. (Level 3)	Mission 2
	2.1 Brainstorming	2.1 Based on the problems, student can propose solutions 2.1.1 Students cannot propose solutions. (Level 0) 2.1.2 Students can propose one solution. (Level 1) 2.1.3 Students can propose two solutions. (Level 2) 2.1.4 Students can propose at least three solutions. (Level 3)	Mission 2
1. Initiation Stage	1.2 Association	1.2 The problems students specify have to correlate with life experience. 1.2.1 Students cannot specify problems correlate with life experience. (Level 0) 1.2.2 Students can specify one problem correlate with life experience. (Level 1) 1.2.3 Students can specify two problems correlate with life experience. (Level 2) 1.2.4 Students can specify at least three problems correlate with life experience. (Level 3)	Question 2, Mission 1
	1.1 Brainstorming	1.1 Based on the question situation, students can specify what problems (e.g., inconvenience) the situation might bring about 1.1.1 Students cannot specify what problems (e.g., inconvenience etc.) the situation might bring about. (Level 0) 1.1.2 Students can specify one problem the situation might bring about. (Level 1) 1.1.3 Students can specify two problems the situation might bring about. (Level 2) 1.1.4 Students can specify at least three problems the situation might bring about. (Level 3)	Question 1, Mission 1

score that is equal to or above *k* (Linacre, 1998). For example, the second Thurstone threshold of a particular item represents the point at which the probability of the participant's getting a score < 1 is equal to the probability of that participant's getting a score ≥ 2. The current study used the second Thurstone threshold for each item as a reference point to compare differences in scientific imagination. Parameters in the Rasch PCM were estimated using the *ConQuest* computer program (Wu, Adams, & Wilson, 2007).

Table 4
Estimates of item difficulty and goodness-of-fit values for scientific imagination.

Dimension	Item	Estimates of item difficulty	*SE*	*OUTFIT MNSQ*	*INFIT MNSQ*
Scientific imagination	1-1	−0.83	0.09	0.80	0.84
	1-2	−0.46	0.09	0.79	0.86
	2-1	−1.40	0.09	1.00	1.02
	2-2	0.77	0.10	0.92	0.99
	2-3	1.63	0.11	0.63	0.92
	3-1	0.30	0.21	**1.44**	**1.52**

Note. The meaning of items 1-1 through 1-3, reference Table 1; the boldface means poor model-data fit.

Data analysis comprised two phases: item revision and validation. During the first phase, the quality of individual items was evaluated to ensure that the characteristics of the four components of scientific imagination were unidimensional. When the data fit the model's expectation, the infit (weighted) and outfit (unweighted) mean square error (MNSQ; range: 0.6–1.4) had the expected value of unity (Bond & Fox, 2007). Next, differential item functioning (DIF; Holland & Wainer, 1993) analyses were conducted between genders and grades. A difference of 0.5 logits in the overall difficulty of items across groups was regarded as substantial DIF (Wang, 2008).

In the validation phase, this study assessed the content validity, structural validity, generalizability, substantive validity, and interpretability (Messick, 1994, 1995a, 1995b; Wolfe & Smith, 2007) of the SIT-Verbal using Rasch PCM analysis.

With regard to content validity, the current study recorded the development of the SIT-Verbal in detail, which included construct definition, items development, expert review, and item revision. With regard to structural validity, the Rasch PCM was used to examine the fit of each item and the unidimensionality of scientific imagination. When data fit the expectation generated by the model, the infit and outfit MNSQ had an expected value of unity. An MNSQ value between 0.6 and 1.4 was set as the criterion for a reasonably good model-data fit (Bond & Fox, 2007). Regarding the substantive aspect of validity, the current hypothesized scoring of each item represented different levels of students' understanding. Theoretically, the hypothesized item hierarchy should match the empirically derived hierarchy. Additionally, we also provided the choosing percentage of each category per item.

With respect to generalizability, DIF analyses were conducted across genders and grades. Also, person separation reliabilities (Schumacker & Smith, 2007) and conditional reliabilities (Raju, Price, Oshima, & Nering, 2007) were calculated for measurement precision. Wright maps are graphical and empirical representations of a construct map, showing how a construct unfolds or evolves in terms of increasingly sophisticated student performances. We used Wright maps in this study to simultaneously show the participants' scientific imagination representation and item difficulties; this allowed the hierarchies of scientific imagination to be evaluated per item.

3. Results of LPs for scientific imagination assessment

3.1. Results of the first phase

Data collection proceeded in two phases. One is for item revision, the other is for validation. In the first phase, responses were obtained from 558 5th and 6th grade elementary students in the southern city of Taiwan. Among these students, it included 51.10% of grade 5th students ($n = 285$), and 48.9% of grade 6th students ($n = 273$); of these, 51.4% ($n = 287$) were male, and 48% ($n = 268$) were female. Three participants did not specify their sex. The age of the students was 11–13 years ($M = 11.86$ years; $SD = 0.72$); 13 students did not specify their age.

The Rasch PCM was used to examine the unidimensionality of scientific imagination. The results revealed that all items in the dimension had acceptable infit and outfit MNSQs (range: 0.6–1.4). DIF analyses were conducted separately for each item. All items in the dimension of scientific imagination did not exhibit substantial DIF for genders and grades.

3.2. Results of the second phase

3.2.1. Content and structural evidence

In the second phase, responses were obtained from 183 5th and 6th grade elementary school students, including 59% of 5th graders ($n = 108$) and 41% of 6th graders ($n = 75$); 50.3% of the participants were male ($n = 92$), and 49.7% were female ($n = 91$). The students were 11–13 years old ($M = 11.73$ years; $SD = 0.67$). The Rasch PCM was used to examine the unidimensionality of scientific imagination. The results revealed that most of the items had acceptable infit and outfit MNSQs (range: 0.6–1.4), with the exception of the item related to conceptualization/organization/formation (Table 4).

3.2.2. Generalizability evidence

DIF analyses were conducted for scientific imagination. All estimates of item difficulty between genders were less than 0.24 logits; the result also showed no substantial DIF was found between grades, with the exception of the item related to conceptualization/organization/formation. The person separation reliability for scientific imagination was .85. The conditional reliabilities for each participant were individually calculated to the precision of measurement. The conditional

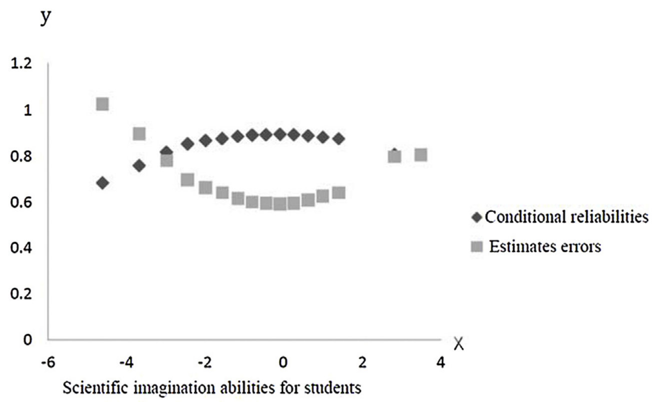

Fig. 2. Conditional reliabilities and estimate errors for abilities.

Table 5
Means and Standard errors for scientific imagination according genders and grades.

Category (*N*)		Scientific imagination
Gender		
Male (92)	*M*	−1.80
	SD	1.79
Female (91)	*M*	−1.32
	SD	1.54
Grade		
5th grader (108)	*M*	−1.73
	SD	1.53
6th grader (75)	*M*	−1.32
	SD	1.87
Total (183)	*M*	−1.56
	SD	1.68

reliabilities and estimation errors for abilities are shown in Fig. 2. The SIT-Verbal was more suitable for measuring students in the mid-level range of understanding (θ range: −1.98 to 1.41 logits; conditional reliability range: .87 to .89) than those functioning at low-level or high-level extremes.

3.2.3. *Interpretability evidence*

Fig. 3 shows the personal measurements and Thurstone threshold item difficulty for scientific imagination. Each "X" on the left side of the figure denotes one person's scientific imagination ability. The digit on the right denotes the item number corresponding to the second Thurstone threshold. Positive values indicate higher levels of scientific imagination ability achieved by that individual; more positive values for an item (i.e., higher levels) indicate that the individual was less likely to have reported scientific imagination performance for that item. The item difficulties of scientific imagination in SIT-Verbal ($M = 0.21$; $SD = 1.27$) were higher than the scientific imagination abilities for 5th and 6th graders' ($M = -1.56$; $SD = 1.68$). Overall, the results confirmed the hierarchy of brainstorming, association, and transformation/elaboration in scientific imagination were confirmed, with the exception of conceptualization/organization/formation. Most of the results from this study were consistent with the findings of Ho et al. (2013).

Table 5 shows differences in scientific imagination between genders and grades. Overall, females ($M = -1.32$; $SD = 1.79$) outperformed males ($M = -1.80$; $SD = 1.53$) in scientific imagination, but the difference was not significant according to *t*-test results ($t = -1.95$, $p = .053$). The results showed no significant difference between 5th ($M = -1.73$; $SD = 1.87$) and 6th graders ($M = -1.32$, $SD = 1.53$) by *t*-test ($t = -1.62$, $p = .108$).

3.2.4. *Substantive evidence*

As expected, our results showed that the higher the ability of the student was, the greater the probability of choosing a higher level. As an example, Fig. 4 shows the Rasch PCM results for the items related to the "brainstorming" component in the initiation stage of Mission 1. The *x*-axis denotes students' brainstorming at different levels (unit: logits); the *y*-axis denotes the probability of each level. The four curves plotted in Fig. 4 correspond to the four levels of understanding for the particular item. Students having an ability of less than −2.66 logits in brainstorming were more likely to choose the Level 0 category. Students having an ability ranging from −2.66 to −0.65 logits, 0.65 to 0.81 logits, and > 0.81 logits were more likely to choose the categories corresponding to Level 1, 2, and 3, respectively. This indicated that the assumption of level setting for each category was well suited to the actual performance of the students.

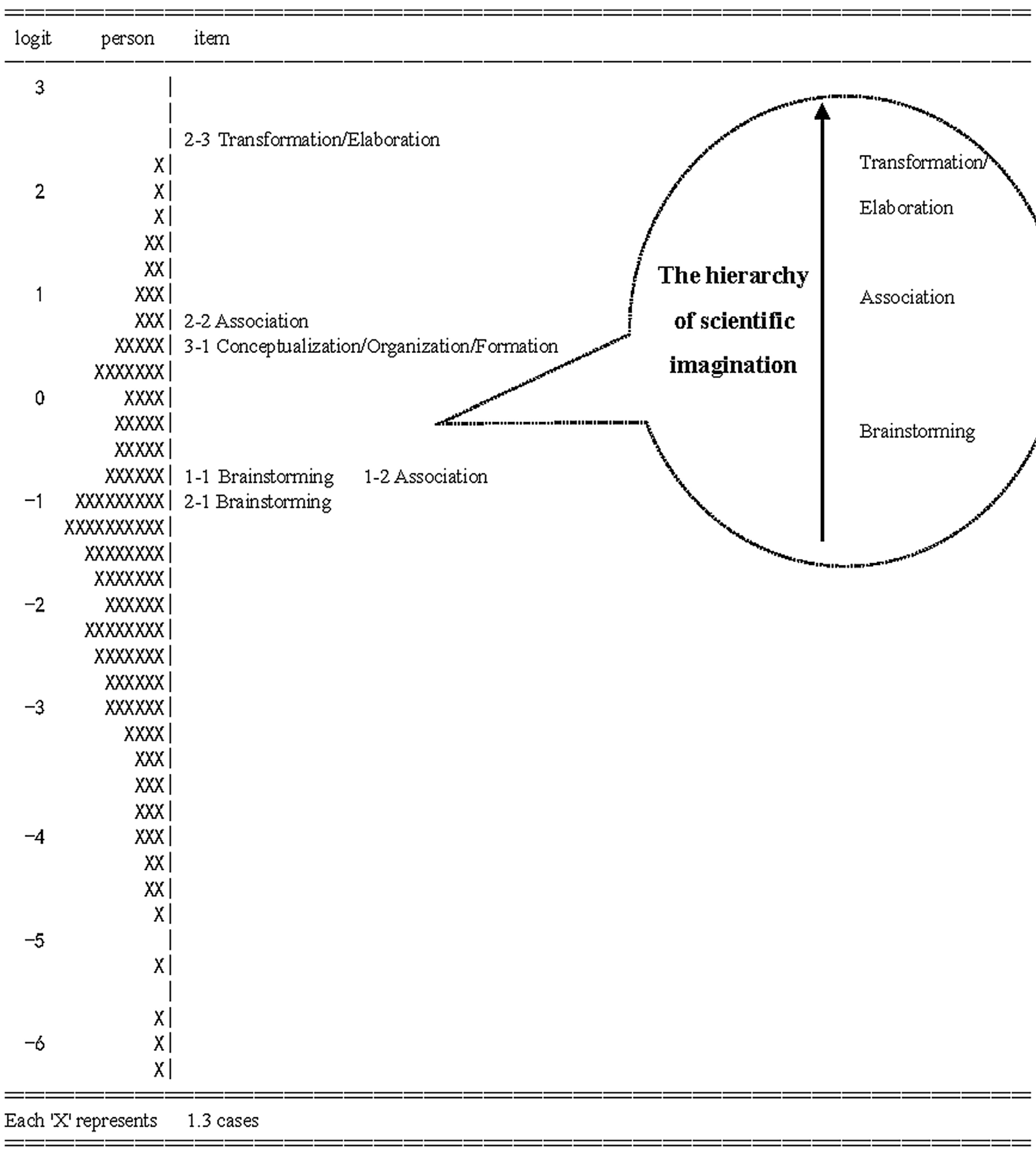

Fig. 3. Personal measures and Thurstone thresholds of items for imagination.

Table 6
Category analysis for Item 1-1.

Level	Score	Count	% of tot	Pt Bis	*t* (*p*)	PV1Avg:1	PV1 SD:1
0	0	52	28.42	−0.65	−11.58(.000)	−3.38	1.54
1	1	65	35.52	−0.16	−2.19(.000)	−1.69	0.94
2	2	44	24.04	0.43	6.50(.439)	−0.39	0.96
3	3	22	12.02	0.57	9.33(.000)	0.77	1.09

Table 6 shows the category analysis for Item 1-1. The percentage for each category was larger than 5%, indicating distractibility in this item. The average estimates for students with different levels (PV1 Avg) showed that the average ability for Level 0 students was the lowest (−3.38 logits), followed by Levels 1, 2, and 3 (−1.69, −0.39, and 0.79 logits, respectively). Thus, these results indicate that the SIT-Verbal was capable of differentiating students' abilities, with the exception of the items related to transformation/elaboration. Level 3 could not be achieved for the transformation/elaboration test item,

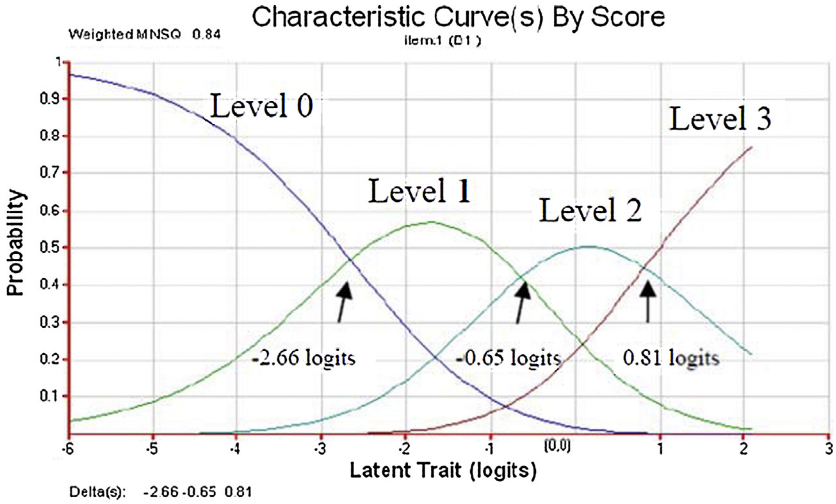

Fig. 4. Probabilities of the four levels of understanding for the student participants, taking item 1-1 as an example.

indicating the need to revise the scoring guide for this item. The revisions may include additional scoring levels for this item as well as different samples.

4. Development and teaching of scientific imagination curriculum

The second part of our research aimed to develop a set of scientific imagination curricula based on the proposed scientific imagination LP as an informal science activity and implemented it in a practical site. In doing so, we attempted to explore the impact of the preliminary curriculum to enrich the holistic curriculum for scientific imagination. Concrete indicators of scientific imagination were used as teaching goals throughout the development of the courses. In the following, we describe the content, impact, and teacher's reflection as those relate to the scientific imagination curriculum in practice.

The curriculum designer was Teacher C, who has 26 years of teaching experience (as of 2013) in elementary schools in southern Taiwan. Teacher C has had students participate in IEYI competitions (representing Taiwan 10 times between 2004 and 2013), and has received 64 awards. This teacher was assisted by a member of our assessment development team for this study.

The curriculum designed by Teacher C was called "Little Inventors" based on the indicators of scientific imagination LP for 5th and 6th graders from July to August in 2013. The curriculum consisted of 12 classes, including units called "Little inventions in daily life," "I am the best in creating," "Doing inventions by myself," and one learning record from different stages in the scientific imagination process. The purpose of the curriculum was to help students to understand that they can invent anytime and anywhere, as well as to make them realize that inventing is not difficult further to motivate their novel ideas and learn science knowledge more from this activity.

The first unit, "Little inventions in daily life," focused on the initiation stage. The unit guided students in their thought processes to consider what problems they might encounter, such as those associated with food, clothing, living, transportation, education, and entertainment. The students were also introduced to the origins of various creations through teacher–student interactions and discussions. The second unit, "I am the best in creating," was based on the dynamic adjustment stage. Teacher C provided students with directions for creating ("stationery box") to encourage associations between the related topic and teacher–student interactions. The students then shared their ideas on how to solve the problem and the functional aspects of their creation. During the process, students were allowed to discuss their creations with the teacher individually. Finally, Teacher C guided the students in writing down their title, motive, purpose, and the function of their creations. The third unit, "Doing inventions by myself," represented the virtual implementation stage. At first, the teacher and students gathered the materials that they had chosen and checked the size of their creations from their drafted design. During the process, Teacher C advised the students on their primary creations and students had to test their creations and write down the creations' mechanisms of operation. Then, students were given advice by other students, which served as a reference point for further modification to their designs. Finally, the teacher evaluated the students' work. A brief introduction and draft of the curriculum were shown in Table 7.

From Teacher C's perspective, when he started to teach the science imagination curriculum, he had to understand students' differences first, implement individual teaching strategies, and provide students with feedback. After teaching each unit, Teacher C recorded his reflections about his teaching. In his reflections, Teacher C indicated that managing teaching time was more difficult with this approach. Numerous class discussions and feedback required adjustments in the teaching schedule. Also, some students were reluctant to participate fully in the class, leading to some frustration for the teacher. Teacher C suggested that teachers should encourage and facilitate student participation in the initiation stage in the future; this would

Table 7
Brief introduction to the curriculum for scientific imagination.

Units	Components	Teaching goals	Procedure	Creations
Little inventions in daily life (three classes)	Brainstorming	1.1 Based on the question situation, students can specify what problems (e.g., inconvenience) the situation might bring about	1. To guide students know inventions more 2. To guide students to appreciate inventions. 3. To introduce principles for inventions 4. To collect unsolved problems in daily life	
	Association	1.2 The problems students specify have to correlate with life experience		
I am the best in creating (three classes)	Brainstorming Association Transformation/Elaboration	2.1 Based on the problems, student can propose solutions 2.2 Students can do with possible solutions and illustrate the physical features and functions of their creations 2.3 Student can reorganize the physical features and functions of creations to solve problems	1. To guide students to think and associate the topics related inventions 2. To discuss with students and guide them to write down the ideas 3. To guide students design creations and draw drafts	 
Doing inventions by myself (six classes)	Conceptualisation/ Organization/Formation	3.1 Students can refine their previous activities to create a prototype for their specific ideas, including issues related to the choice of materials, techniques for assembling parts, and the means of creating design diagrams and final drafts from the initial diagrams	1. To guide students to do inventions 2. To guide students to write their reports for inventions 3. To guide students to test and modify their creations 4. Students share with their ideas to other students	

Resources designed by Teacher C himself.

allow the teacher to collect additional information to guide the students in the discussion phase of the dynamic adjustment stage. When entering the virtual implementation stage, students may not prepare sufficient materials to practice. Thus, teacher oversight may be required to facilitate the students' creativity.

Additionally, when teachers interact with students, the student's response cannot always be predicted; thus, teachers have to consider in advance how they will reach teaching goals efficiently and then employ various teaching strategies to guide students during the process of teaching. For example, in the dynamic adjustment stage, Teacher C thought that several of the students proposed some rather far-reaching ideas; however, he noted that it is important that teachers refrain from negating students' ideas. He felt that instead, teachers should encourage students' willingness to share. Teachers should not be afraid that these students might sidetrack the class objectives. Teachers can continuously modify and replenish ideas. For teachers, it is a difficult but crucial process to ensure that the teaching process runs smoothly. However, students' feedback from their peers is an important part of the innovation process.

Indeed, it is a challenge for teachers to teach extra science innovation activities, particularly when these activities include various factors and proactive student involvement in moving the process forward. Therefore, teachers should have sufficient content knowledge, pedagogical strategies and be open to accepting others' opinions, as well the ability to offer students structured suggestions to open various innovation pathways for the students. Regarding the curriculum developed in this study, a number of modifications are needed. Nevertheless, this kind of informal scientific activity enhances not only teachers' effectiveness but also students' ability to create. Future suggestions for curriculum design are provided.

5. Discussion and conclusion

The current study aimed to explore the high-level thinking skills entailed in scientific imagination based on the concept of LPs. We attempted to build a LP for scientific imagination to align curriculum, instruction, and assessment through the BAS and clarify the core ideas and the developmental path of the process of scientific imagination so as to promote scientific innovation.

Regarding development of assessment based on proposed LP, first, the results in terms of content and structural validity indicated that the items included in the SIT-Verbal had good model–data fit, with the exception of the conceptualization/organization/formation step. Further analysis indicated that the scoring of this item may need to be revised to include additional levels instead of limiting the scoring to four levels. The item for conceptualization/organization/formation required the study participants to draw one "new invention" that they thought would solve the problem; students were asked to include the choice of materials, techniques for assembling the parts, design diagrams, and final drafts from the initial diagrams. We speculated that the number of levels of understanding for this particular item should not have been limited to four levels (scoring levels 0 to 3). The scoring might be needed to include additional levels to differentiate students' understanding further to verify the LP of scientific imagination. For example, a score of 3.1.2 may indicate that "Students can roughly propose the idea of the invention they will create through speaking or writing." In this case, "roughly" may be too abstract for those rating the participants' responses.

Moreover, Rasch fit statistics (mean squares and t-statistics) are highly susceptible to sample size variation for dichotomously scored rating data. But Smith, Rush, Fallowfield, Velikova, and Sharpe (2008) indicated that *t*-statistics were highly sensitive to sample size, whereas mean square statistics remained relatively stable for polytomous data. Although there is not a large sample size in the second phase, we can certainly perform useful exploratory work using Rasch analysis with a small sample (Linacre, 1994). Besides, the reason that this item did not fit the model may be the requirement of two abilities: drawing and describing. After further discussion, we decided to include this item due to its importance to scientific imagination; however, additional scoring levels and larger sample size may be included in future versions of the test. In future studies, the perspective offered by Amabile (1983, 1996) could be also used to provide a better indicator of creativity by assigning *creativity* and *technical goodness* scores to the prototypes or drafts.

Regarding generalizability, the person separation reliability of scientific imagination was .85. Our test results indicated that the SIT-Verbal was more suitable for mid-level students than for those functioning at an extremely low or high level. In practice, this outcome was consistent with normal class groupings. Thus, the SIT-Verbal is an adequate tool to assess students' abilities of scientific imagination, especially for average students.

In terms of interpretability, the results showed that the hierarchy of four components did not include the conceptualization/organization/formation in its current form. In the future, the scoring guide for this component will need to be revised and verified using different samples for this item. The current study hypothesized that the scientific imagination process should be cyclical. However, the results only verified the stages in the scientific imagination process. Future studies in this area should verify the relationships among constructs for the scientific imagination process, using structural equation modeling (SEM).

Substantive evidence showed that the higher the ability of the student was, the greater the probability was of their choosing higher levels in their response. This indicated that the assumption of a level setting in each category coincided with the real performance of the student. If a particular scoring level could not be achieved by the students as a whole, then this suggested that additional categories were needed between the highest level achieved and the maximum achievable level to provide more information. For example, if no student could reach Level 3 for a particular item, then the guide for scoring this level should include additional categories between Levels 2 and 3 to provide more information.

No significant difference in scientific imagination was evident between genders or grades. However, various daily life experiences and cognitive development for different ages should be considered. The development of the imagination begins in childhood and continues into adulthood. As a person grows, the imagination undergoes various transformations (Eckhoff & Urbach, 2008; Vygotsky, 1930/2004). Thus, the development and verification of LPs must to be revised continuously throughout the development process to accommodate the age of the learner (Shea & Duncan, 2013; Songer et al., 2009). Future studies should expand the grades sampled to include 3rd to 6th graders so as to verify the SIT-Verbal as a reference for scientific imagination curriculum development and future innovation studies.

Furthermore, the first trial to develop a curriculum based on the proposed scientific imagination LP was also conducted. The curriculum of scientific invention was difficult for teachers to control in practice. The observations from Teacher C's reflections indicated the need to revise connections among curriculum, instruction, and assessment in the future. In this case, checklists may help teachers to implement and assess scientific imagination instruction and students' LP in practice. Developing a set of formal curriculum infusing scientific imagination and examining the experimental effect in the future would be needed to better understand the core ideas and the developmental path of the scientific imagination process as well as align curriculum, instruction, and assessment through LP. This approach could potentially enhance teachers' effectiveness as well as provide a wealth of information to better understand the relationship of LP, scientific imagination and innovation.

In conclusion, a review of past literature revealed that much of the research on LPs is in its infancy, and current models are largely conjectural owing to gaps in the research base (Duncan & Rivet, 2013). Nevertheless, the current study offered a productive starting place for developing standards, curricula, and assessments in scientific imagination. It is notable that the LPs for scientific imagination in the study are built in the 5th and 6th grade elementary school level. The result may not be able to be inferred to other levels. Furthermore, the LPs for scientific imagination might be adjusted regarding age, gender, professional backgrounds, etc. Building LPs is an iterative and developmental process, and an ongoing research is required. In the future, we plan to revise the scoring guide for the transformation/elaboration and conceptualization/organization/formation components, as well as to verify the LP process in different samples. Additionally, the SIT-Verbal will be used in the curriculum to align curriculum, instruction, and assessment on future LPs. Reflections and feedback from teachers and students at each stage of the scientific imagination process will further enhance teacher effectiveness and students' ability to conceptualize and create.

Acknowledgement

This research was supported by grants from the Ministry of Science and Technology of Taiwan (National Science Council of Taiwan) under contract no. NSC100-2511-S-110-002-MY3.

References

Adadan, E., Trundle, K. C., & Irving, K. E. (2010). Exploring grade 11 students' conceptual pathways of the particulate nature of matter in the context of multirepresentational instruction. *Journal of Research in Science Teaching*, *47*, 1004–1035. http://dx.doi.org/10.1002/tea.20366

Adams, J. (2004). The imagination and social life. *Qualitative Sociology*, *27*, 277–297. http://dx.doi.org/10.1023/B:QUAS.0000037619.28845.ef

Alonzo, A., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, *93*, 389–421. http://dx.doi.org/10.1002/sce.20303

Amabile, T. M. (1983, 1996). *Creativity in context: Update to the social psychology of creativity*. Boulder, CO: Westview Press.

Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, *94*, 765–793. http://dx.doi.org/10.1002/sce.20402

Black, P., Wilson, M., & Yao, S. Y. (2011). Road maps for learning: A guide to the navigation of learning progressions. *Measurement*, *9*, 71–123. http://dx.doi.org/10.1080/15366367.2011.591654

Bond, T., & Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: LEA.

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment*, *11*, 33–63. http://dx.doi.org/10.1207/s15326977ea1101_2

Campbell, K. H. S., Mcwhir, J., Ritchie, W. A., & Wilmut, I. (1996). Sheep cloned by nuclear transfer from a cultured cell line. *Nature*, *380*, 64–66. http://dx.doi.org/10.1038/380064a0

Church, E. B. (2006). Let's invent. *Scholastic Parent & Child*, *13*, 28–35.

Claesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The perspectives on chemists. *Science Education*, *93*, 56–85. http://dx.doi.org/10.1002/sce.20292

Cruz, H. D., & Smedt, J. D. (2010). Science as structured imagination. *Journal of Creative Behavior*, *44*, 29–44. http://dx.doi.org/10.1002/j.2162-6057.2010.tb01324.x

Duncan, R. G. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching*, *46*, 606–609. http://dx.doi.org/10.1002/tea.20316

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal for Research in Science Teaching*, *46*, 606–609. http://dx.doi.org/10.1002/tea.20316

Duncan, R. G., & Rivet, A. E. (2013). Science learning progressions. *Science*, *339*, 396–397. http://dx.doi.org/10.1126/science.1228692

Duncan, R. G., Rogat, A. D., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th–10th grades. *Journal of Research in Science Teaching*, *46*, 655–674. http://dx.doi.org/10.1002/tea.20312

Duncan, R. G., & Tseng, K. A. (2011). Designing project-based instruction to foster generative and mechanistic understandings in genetics. *Science Education*, *95*, 21–56. http://dx.doi.org/10.1002/sce.20407

Eckhoff, A., & Urbach, J. (2008). Understanding imaginative thinking during childhood: Sociocultural conceptions of creativity and imaginative thought. *Early Childhood Education Journal*, *36*, 179–185. http://dx.doi.org/10.1007/s10643-008-0261-4

Freidenreich, H. B., Duncan, R. G., & Shea, N. (2011). Exploring middle school students' understanding of three conceptual models in genetics. *International Journal of Science Education*, *33*, 2323–2349. http://dx.doi.org/10.1080/09500693.2010.536997

Fulmer, G. W., Liang, L. L., & Liu, X. (2014). Applying a force and motion learning progression over an extended time span using the force concept inventory. *International Journal of Science Education*, http://dx.doi.org/10.1080/09500693.2014.939120

Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, *94*, 258–281. http://dx.doi.org/10.1002/sce.20368

Grant, E. (2004). Scientific imagination in the middle ages. *Perspectives on Science*, *12*, 394–423. http://dx.doi.org/10.1162/1063614042776021

Ho, H. C., Wang, C. C., & Cheng, Y. Y. (2013). Analysis of the scientific imagination process. *Thinking Skills and Creativity*, *10*, 68–78.

Holland, P. W., & Wainer, H. (Eds.). (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.

Johnson, P., & Tymms, P. (2011). The emergence of a learning progression in middle school chemistry. *Journal of Research in Science Teaching*, *48*, 849–877. http://dx.doi.org/10.1002/tea.20433

Lee, H. S., & Liu, O. L. (2010). Assessing learning progression of energy concepts across middle school grades: The knowledge integration perspective. *Science Education*, *94*, 665–688. http://dx.doi.org/10.1002/sce.20382

Linacre, L. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*, 328.

Linacre, J. M. (1998). Thurstone thresholds and the Rasch model. *Rasch Measurement Transactions*, *12*, 634–635.

Liu, X. F., & Lesniak, K. (2006). Progression in children's understanding of the matter concept from elementary to high school. *Journal of Research in Science Teaching*, *43*, 320–347. http://dx.doi.org/10.1002/tea.20114

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.

McCormack, A. J. (2010). Imagine and invent: Create a great future. *Journal of College Science Teaching*, *22*, 14–15.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, *23*, 13–23. http://dx.doi.org/10.3102/0013189X023002013

Messick, S. (1995a). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice*, *14*, 5–8. http://dx.doi.org/10.1111/j.1745-3992.1995.tb00881.x

Messick, S. (1995b). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749. http://dx.doi.org/10.1037/0003-066X.50.9.741

Mey, T. D. (2006). Imagination's grip on science. *Metaphilosophy*, *37*, 222–239. http://dx.doi.org/10.1111/j.1467-9973.2006.00432.x

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, *46*, 675–698. http://dx.doi.org/10.1002/tea.20314

Mountain, V. (2007). Educational contexts for the development of children's spirituality: Exploring the use of imagination. *International Journal of Children's Spirituality*, *12*, 191–205. http://dx.doi.org/10.1080/13644360701467535

National Research Council. (2006). Systems for state science assessment. Committee on Test Design for K-12 Science Achievement. In M. Wilson, & M. Bertenthal (Eds.), *Board on testing and assessment, Center for Education. Division of Behavioral and Social Sciences and Education*. Washington, DC: National Academy Press.

National Research Council. (2007). In R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.), *Taking science to school: Learning and Teaching Science in Grages K-8. Committee on Science Learning, Kindergarten through Eighth Grade*. Washington, DC: National Academy Press.

Osborn, A. (1953). *Applied imagination*. New York: Charles Scribner's Sons.

Pendry, J. B., Schurig, D., & Smith, D. R. (2006). Controlling electromagnetic fields. *Science*, *312*, 1780–1782. http://dx.doi.org/10.1126/science.1125907

Plummer, J. D., & Krajcik, J. (2010). Building a learning progression for celestial motion: Elementary levels from an earth-based perspective. *Journal of Research in Science Teaching*, *47*, 768–787. http://dx.doi.org/10.1002/tea.20355

Raju, N. S., Price, L. R., Oshima, T. C., & Nering, M. L. (2007). Standardized conditional SEM: A case for conditional reliability. *Applied Psychological Measurement*, *31*, 169–180. http://dx.doi.org/10.1177/0146621606291569

Robinson, A. (2010). Chemistry's visual origins-Vivid imagination was key to unlocking the secrets of molecular structure in the nineteenth century [Review of the book Image and Reality: Kekulé, Kopp, and the Scientific Imagination, by A. Rocke]. *Nature*, *465*, 36. http://dx.doi.org/10.1038/465036a

Schumacker, R. E., & Smith, E. V., Jr. (2007). Reliability: A Rasch perspective. *Educational and Psychological Measurement*, *67*, 394–409.

Schwarz, C. V., Reiser, B. J., Davis, E. A., & Kenyon, L. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*, 632–654. http://dx.doi.org/10.1002/tea.20311

Shea, N. A., & Duncan, R. G. (2013). From theory to data: The process of refining learning progressions. *Journal of the Learning Sciences*, *22*, 7–32. http://dx.doi.org/10.1080/10508406.2012.691924

Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*, *8*, 1–11. http://dx.doi.org/10.1186/1471-2288-8-33

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic–molecular theory. *Measurement*, *14*, 1–98.

Songer, N. B., & Gotwals, A. W. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching*, *49*, 141–165. http://dx.doi.org/10.1002/tea.20454

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*, 610–631. http://dx.doi.org/10.1002/tea.20313

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*, *46*, 699–715. http://dx.doi.org/10.1002/tea.20308

Stevens, S. Y., Delgado, C., & Krajcik, J. S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, *47*, 687–715. http://dx.doi.org/10.1002/tea.20324

The Next Generation Science Standards. (2013, April). *Appendix H – Understanding the scientific enterprise: The nature of science in the next generation science standards*. Retrieved from http://www.nextgenscience.org/sites/ngss/files/Appendix%20H%20-%20The%20Nature%20of%20Science%20in%20the%20Next%20Generation%20Science%20Standards%204.15.13.pdf

Vygotsky, S. L. (1930/2004). Imagination and creativity in childhood. *Journal of Russian and East European Psychology*, *42*, 7–97.

Wang, W. C. (2008). Assessment of differential item functioning. *Journal of Applied Measurement*, *9*, 1–22.

Wang, C. C., Cheng, Y. Y., Liu, K. S., & Ho, H. C. (2011). Validation of the multidimensional sense of humor scale using Rasch analyses. *Psychological Testing*, *58*, 691–713 (in Chinese).

Wang, C. C., Ho, H. C., Cheng, C. L., & Cheng, Y. Y. (2014). Application of the Rasch model to the measurement of creativity: The Creative Achievement Questionnaire. *Creativity Research Journal*, *26*, 62–71. http://dx.doi.org/10.1080/10400419.2013.843347

Wang, C. C., Ho, H. C., Wu, J. J., & Cheng, Y. Y. (2014). Development of the scientific imagination model: A concept-mapping perspective. *Thinking Skills and Creativity*, *13*, 106–119. http://dx.doi.org/10.1016/j.tsc.2014.04.001

Ward, T. B. (1994). Structured imagination: The role of category structure in exemplar generation. *Cognitive Psychology*, *27*, 1–40. http://dx.doi.org/10.1006/cogp.1994.1010

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Wilson, M. (2009). Measuring processions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching*, *46*, 716–730. http://dx.doi.org/10.1002/tea.20318

Wolfe, E. W., & Smith, E. V., Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II. Validation activities. *Journal of Applied Measurement*, *8*, 204–233.

Wu, M. L., Adams, R. J., & Wilson, M. R. (2007). *ConQuest [Computer software and manual]*. Camberwell, Victoria, Australia: Australian Council for Educational Research.

Zabriskie, B. (2004). Imagination as laboratory. *Journal of Analytical Psychology*, *49*, 235–242. http://dx.doi.org/10.1111/j.1465-5922.2004.00455.x

Zhao, M., Hoeffler, S., & Dahl, D. W. (2009). The role of imagination-focused visualization on new product evaluation. *Journal of Marketing Research*, *46*, 46–55. http://dx.doi.org/10.1509/jmkr.46.1.46