南台人文社會學報 2013 年 11 月
第十期 頁 109-135

幼兒科學創造力實作評量類推性之研究

侯雅齡*

摘要

本研究在探討幼兒科學創造力實作評量分數的類推性。幼兒科學創造力的評量乃嵌入於科學創意教學中進行評估，讓幼兒在具創造經驗與創造過程的活動中發展出有意義的學習並萌發創意。實作評量包含實作任務、反應形式與評分，其中，實作任務為五個動態的簡易動手做物理活動，每一個動手做活動都透過不同的反應形式與使用不同的評量工具，來取得幼兒在科學創造力特質、心流（flow）與成果三向度資料。本研究使用的評量工具包含：幼兒科學創造力評量表、幼兒心流經驗量表與科學產品效能評分表。在以類推性理論進行量化的分析的結果，科學創造力實作評量在五個任務作業、三個評分向度的 P(person)×D(dimension)×T(task)交叉設計的類推性係數（ρ^2）為.811，擁有良好的信度訊息。整體而言，幼兒科學創造力實作評量模式，有別於一般採紙筆測驗的創造力評量傳統，讓教學與評量更為貼近，符合幼兒創造力評量的趨勢，值得作進一步的推廣與應用。

關鍵詞：科學創造力、類推性理論、實作評量

* 侯雅齡，國立屏東教育大學特殊教育學系副教授
電子信箱：yalingho@mail.npue.edu.tw
收稿日期：2013 年 03 月 18 日，修改日期：2013 年 09 月 05 日，接受日期：2013 年 11 月 22 日

STUST Journal of Humanities and Social Sciences, November 2013
No. 10 pp.109-135

The Generalizability of Preschool Children's Science Creativity Performance Assessment

Ya-Ling Hou[*]

Abstract

The purpose of this study was to assess the reliability of an instrument for assessment of preschool children's science creativity performance developed by the researcher. To better understand children's creativity, we provided children with activities requiring imagination, and then observed their creativity in terms of both process and outcome. Different curriculum-embedded performance tasks, response formats and scoring were implemented in a three-stage teaching process for the study. The preschoolers' scientific creativity was assessed in five hands-on science activities in physics, with three assessment dimensions (trait, flow and product) applied for each activity. The Preschool Children's Science Creativity Scale, Preschool Children's Science Flow Experience Scale, and Creativity Products Evaluation Criteria were used. The generalizability coefficient (ρ^2) of the five performance tasks and three dimensions achieved .81 under the cross design of P(person)× D(dimension)× T(task). The results of the performance assessment and the findings of this study offer positive suggestions for the public to pay more attention to the potential of children for creativity in scientific activities.

Keywords: Science creativity, Performance assessment, Generalizability theory

[*] Ya-Ling Hou, Associate Professor, Department of Special Education, National Pingtung University of Education
E-mail: yalingho@mail.npue.edu.tw
Manuscript: Mar. 18, 2013; Modified: Sept. 05, 2013; Accepted: Nov. 22, 2013

Introduction

The study of creativity has been a topic of intense research of psychologists and educators. Researchers have dealt with the nature of creativity and paid attention to the questions such as, "what is creativity?" or "where is creativity?" Especially after Guilford made an appeal to the public on 1950, there came out lots of related researches continuously. Except for the 4 P's denoting the four generally accepted facets to creativity, namely Person, Process, Product and Place, after the eighties, Simonton (1984), Gardner (1993), Sternberg & Lubart (1995), Amabile (1995), and Csikszentmihalyi (1996) proposed confluence approach to creativity. They believed that multiple components must converge in order for creativity to occur. Right until today, the confluence approach has become the mainstream to creativity studies (Wu, 2002; Sternberg, 1999). After the Ministry of Education published the White Paper on Creative Education on 2002, the government made efforts to push forward the researches on creativity. However, it was difficult to explore children's creativity through scientific approach. Constricted to the developmental factors, children were limited on expression, logical and targeted thinking. Their knowledge and experiences were insufficient as well, not to mention the invention of novel and effective products. Unfortunately, those on the way used in standard assessment. Therefore, it was difficult to apply standardized assessment to children, and the related researches were not easily to come out.

Hou (2009) had developed preschool children's science creativity performance assessment. Performances in the preschoolers' scientific creativity assessment included 5 hands-on science activities in Physics and 3 assessment dimensions (trait, flow and product) in each activity. In this study was to study the impact of tasks, assessment dimensions on reliability of preschool children's science creativity performance assessment.

Children's Scientific Creativity

Runco (2006) described children's creativity in two concepts – stages and domains. First, the creativity of preschool children was different from that of school-aged children, which in turn differed from that of adolescents and adults. Therefore, if pulled out from the premise of stage, we couldn't really understand creativity, and even misunderstood that children do not have creativity. The creativity of adults often leads to some products – a work of art, a solution to problem, but children's creativity may not produce a tangible product or result. It was just an imaginative play or exploration (Russ, 1994; Smoluch, 1992). This was the course that children discovered who they are and what is acceptable in their family, school, peer-group, and culture. Young children did things that satisfy two requirements of creativity – originality and usefulness. Second, domain of creativity performance, had been recognized in studies of creativity (Csikszentmihalyi, 1996), for instance, a child who had creative potential in the musical domain may be quite dissimilar to the child who had creative potential in the mathematical, verbal, athletic, or some other domains. By big C and little C concept of Csikszentmihalyi, Craft (2002) deemed that the children's creativity is little C that is to say the contributory creation behavior of daily life. Because the characteristic of domain differences had already been represented on children, considering the concept of domain differences was essential in study the creativity of children. In this study, the researcher investigated creativity in children's scientific domain.

The NACCCE (National Advisory Committee on Creativity and Culture in Education, 1999, p. 30) suggested that creativity is an "imaginative activity fashioned so as to produce outcomes that are both original and of value".

This definition included four key points: thinking or behaving imaginatively, imaginative activity being purposeful, imaginative activity that generates something original and the outcome being of value to the original objective. Regarding children as the subject, Feasey (2005, pp. 3-5) interpreted these four key points as the following:

1. **Thinking and behaving imaginatively:** Imaginative activity was the process of generating something original. In playing, children could explore different possibilities and look for the relationship between each other. The teacher must be the activator to encourage children to share ideas, pose questions and problems, and create more and more imagination. Nickerson (1999, p.410) suggested that children maintaining their curiosity into adulthood depend to a large degree on the extent to which it is encouraged or inhibited in early life.
2. **Imaginative activity being purposeful:** Creativity was purposeful, most of each creative thought and action had a reason (purpose) behind it. As to children, the idea of creativity was very extensive, including to work towards producing something, an idea to share, a question to answer, a problem, an artifact, or indeed finding a way of working etc..
3. **Being original:** If we define creativity as being original to the individual, then this allows for everyone to have creative potential. In terms of primary science, two important categories were described. One was that a person's work may be original in relation to their own previous work and output. The other was that it may be original in relation to their peer group. For instance, a child said 'put petrol into the wheels, the car moves!' The idea was not correct, but it was the result of the child using personal observations and knowledge to be purposeful in his thinking. The fact that it was not entirely correct should not diminish the creative

sense of the individual.

4. **The outcome being of value to the original objective:** The outcome of imaginative activity could only be called creative if it is of value in relation to the task at hand. Creativity was possible in all areas of human activity and all young people and adults had creative capacities. Developing these capacities involved a balance between teaching skills and understanding, and giving children time and space to take risks.

To sum up, children's exploration process were just like scientists. We should understand this kind of characteristic, so that we could probe into children's science creativity from their perspectives. From the above mentioned definition of children's science creativity, we could know that children's creativity wouldn't come from nowhere. If we want to observe children's creativity, we should provide children with imaginative activities or experiences, and then to observe or comprehend their creativity through the process and outcome.

Children's Science Performance Assessment

Performance assessment is an alternative assessment. Students under a certain specific situation processed and accomplished a task, and then experts assessed the process (for instance, manipulating of the tools) and results (for instance, the products or outcome reports) (Wiggins, 1993, 1998). Doran, Boorman, Chan and Hejaily (1992) pointed out the definite mismatch between hands-on, inquiry-based science curriculum and old traditional paper-and-pencil methods for testing student knowledge. They indicated the importance of performance assessment in understanding students' scientific learning effect.

Children's creative characteristic was not easily understood by written tests. As a matter of fact, it was spread in the process of learning. For children, the boundary between the course learning and product assessment was indefinite. Gardner (1993) suggested that assessment should proceed in the special situation related to its field, and need to offer proper material and experience to children. Brooke and Solomon (1998) said in their study, having fun in playing brings significant exploration. If students concentrated on playing, they will be curious about it. They may be passive in initial stage, but will turn to be active gradually. And then they will explore, improve intentionally, so playing will transfer to a significant investigation event rally. Craft (2002) also emphasized the relation between playing and the development of creativity; a creative teacher should put the playing elements into teaching. Therefore, the researcher planned to offer children animated and active learning surroundings to let students manipulate, explore and learn from the materials freely to manifest their creativity. During students' playing, teachers observed students' flow and all kinds of creative behavior as the evidence to understand children's creativity. In scientific domain, physics knowledge was to probe into the objects that can be seen in the 'external world'. For children, physics activity was to 'move' the objects, so it interested children and provided the opportunity to experience operation and observation (Kamii & DeVries, 1978). Therefore, the researcher provided children experiences to create and activities of creation process by hands-on, inquiry-based physics curriculum and hoped to motivate children to become involved in exploring the process and result, to observe and understand their scientific creativity performance.

Brown and Shavelson (1996) emphasized the relation between performance assessment and hands-on science curriculum, extended the general components of performance assessment – performance task and scoring system to three elements – performance task, response format and scoring system (Figure 1), expected teachers to construct the performance assessment system on these three elements, and the response format connected compactly to the teaching and assessment. Each of these components is listed in detail below.

1. **Performance Task:** Students were given an invitation to solve a problem or conduct an investigation and they were provided concrete materials and condition to do so, expecting to understand students' ability by their performance.
2. **Response Format:** There were many different response formats in result presenting. Students were asked to use objects, investigate and observe the result, and even to make one product. The response formats were different depending on the distinct level of performance limitation. Relative to the limited response format which was usually more structural, needed less time to measure, but students could not obtain the entire amount of information and original skill, the less limited response format gave liberty and let students prove, reorganize and display individual original idea, but it took time and was difficult to assess, and was even hard to extrapolate to other performances.
3. **Scoring System:** Clear scoring criterion was an important element in performance assessment. If there is not a clear standard, you could not lead students to the achievement; and of course could not assess their performance. A clear scoring standard was helpful to explain the expectation of task, the goal and standard of learning. On the other hand, it also enhanced reliability and the validity of assessment. There were two scoring modes in the process and result assessment. One was scoring rubrics / rating scale, the other was checklists. We could choose a scoring system according to different response formats.

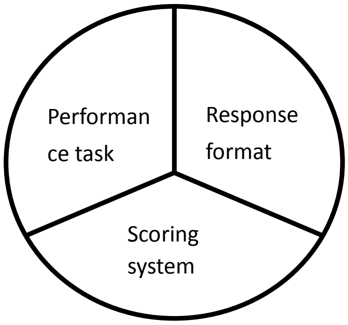

Figure 1 Components of a performance assessment

In choosing or constructing the performance assessment, there were some important characteristics we should concern about. Brown and Shavelson (1996) suggested that in addition to reliability and validity, we could not neglect utility and practicality.

1. Reliability was consistency of results. Performance assessment was a judging assessment; it didn't have clear evidence as in a traditional test. (Dunbar, Koretz & Hoover, 1991) The psychometric analysis was to understand the consistency between different raters, tasks and occasions. Due to numerous and complicated sources of error, the reliability assessment in Classical Test Theory was limited; each time could only take one source of error to estimate the consistency of score. Shavelson and Webb (1991), and Brennan (1983, 2001) indicated generalizability theory to assess the reliability of performance assessment, using analysis of variance to discriminate different sources of error in assessment to estimate the proportion of each variance. It was very helpful to analyze reliability of performance assessment. Therefore, this study planned to

adopt generalizability theory to understand the error that due to measuring tools and to design a more effective performance assessment.

2. Validity of performance assessment explained the appropriateness of assessment result. There were three kinds of validity, including content validity, exchangeability of assessment methods and construct validity. Messick (1989) indicated to consider the results of assessment explanation and application, therefore complete consideration to the content, construct, criteria and results explanation could provide good evidence for validity of test, but requirement of four validities at the same time was unnecessary and impractical, it should depend on the intention of the assessment (Linn & Gronlund, 2000).
3. Utility was also a key element to choose and develop an assessment. Utility could be viewed from the result of assessment: Is the result of assessment in good explanation and application? Does the teacher make a wiser educational strategy by understanding individual student or group's performance? Does the student know his own performance and how to promote?
4. Practicality required teacher's time, energy and even money. For children, the process of hands-on was joyful, but for the teacher, it needed significant efforts on preparation, assessment and observation of the activity.

In sum, children's creativity was first considered on stage and domain. Children's creativity was little C creativity. The scientific creativity performance could be seen by the question solving process of hands-on activities. To understand children's creativity we needed to stimulate creative ability by creative instruction and to give performance tasks for the assessment of creative performance in different stages.

Research Approach

Participants

1. This study took 10 six-year-old preschool children as pilot study samples and 22 six-year-old preschool children as formal samples. They were participants of this curriculum- embedded performance tasks.
2. The teacher of this study was a master of physics with abundant preschool education experiences. After adequate discussion on educational belief with the researcher, teacher carried out action research in the pilot study to insure the execution of this program.
3. The raters of this study were the teachers of those participants' preschool children. In the pilot study, the researcher did the inter-rater reliability.

Training for the Raters

Concerning the raters should have certain degree of familiarity with the children, we chose the mentors of the children to be the raters. Before the units being taught, the raters should familiar with the course design and the lesson plan, and then have discussion with the researcher for about forty to fifty minutes in order to fully understand the meaning of the rubrics. Both the researcher and the raters had their tentative evaluation as a warm-up before the second pilot study, and immediately discussed the inconsistency and calculated the consistency coefficient. The coefficients of the two tentative evaluations were 0.76 and 0.88. Due to the concrete and detailed illustration of the scoring rubrics, we only had to have clear communication and full discussion so that we could have high degree of the consistency coefficient.

Performance Task－Hands-on Physical Curriculum

This course was constructed on creative science mode of Feasey (2005), taking simple physical activities as performance tasks.　For gathering more creative data, different goals assessed by different response formats were set up in every stage of dynamic teaching activities.

The table 1 is this performance task course and teaching lesson plan.

Table 1

Performance task and lesson plan

Stage	Teaching activity (Task)	Response Format	Scoring	Goal
Explore	1. Promote learning motivation: pictures books or scientific finished products. 2. Hands-on activity: scientific challenging question (oral), group sessions. 3. Presentation and discussion	1. Hands-on. 2. Group sessions. 3. Having fun in playing. 4. Presentation.	1. Preschool Children’s Science Creativity Scale. Preschool 2. Children’s Science Flow Experience Scale.	1. Defining the problem. 2. Considering prior experience and knowledge to solve the problem. 3. Discuss, inspect and help each other. 4. Think another ways or ideas to solve the problem. 5. By appropriate scaffolding and question design, obtain correct scientific knowledge and complete the products with internalization. 6. Think to ask for help or not, and what kind of assistance can inspire to solve further problem. 7. Having fun.
Deepen	1. Scientific challenging question (work sheet). 2. Design different Challenging Games, each checkpoint gives one scientific question (variables investigation). 3.Presentation and discussion	Challenging Games. Consider other variables that may affected the finished products. Individual opinion presentation.	1. Children’s scientific knowledge work sheet.* 2. Children’s Science Flow Experience Scale.	
Achieve-ment	1. Situational arrangement, products presentation to promote more improvement and thinking.	1. Complete the finished product. 2. Operate the finished product.	1. Creativity Products Evaluation Criteria 2. Children’s Science Flow Experience Scale.	

Note: The duration of each stage takes one class.

*In formal study, it is teaching activity rather than assessment.

The characteristic of this course included:

1. Emphasized children's motive inspiring.
2. Provided challenging question to link children's old experience and carry out purposeful thinking.
3. The arrangement of materials inspired creative characteristics.
4. Provided abundant time to let children have fun.
5. In a deeper stage, provided appropriate scaffolding in work sheet to encourage children to explore in system.
6. Encouraged children to discuss inspect and help each other.

In this course, there were five hands-on physical activities as active performance tasks. The design of performance task took account of utility and practicality of teaching, the limitation from children's physiological development and curiosity promotion. The scientific hands-on by using common and cheap materials in our daily lives was easy to produce and succeed. Furthermore, in order to conform to the construct of generalizability study, the course took dynamic activities in which children could easily control the contents. "Shooting arrow", "air gun", "paper cup propeller", "wire-walker" and "jumping bean" were the five contents. The utility and practicality were assessed by a 5-point Likert scale questionnaire after the course observation by five kindergarten teachers, and the average scores were 4.6 and 4.8, it proved this course was well designed.

Scales

In considering the learning process and result in assessment, understanding of children's creative characteristics, motive and performance, the researcher used the following four scoring systems.

1. The Preschool Children's Science Creativity Scale

In creative science mode of Feasey (2005), sensitivity, fluency, flexibility and originality could be seen in the process of children's problem solving. Therefore, researcher put scientific challenging questions in hands-on activities to let children solve the problem by group sessions and self-regulation, in order to assess their creative characteristics.

There were 10 questions in the scale; including flexibility, fluency, sensitivity, originality and cooperation, five dimensions in each there were two questions, in the form of 3-point scales (Linn & Gronlund, 2000; Stiggins, 1994). After judging the degree of each child's behavior or performance, the teacher gave a 0, 1 or 2. The scoring rubrics of each question (to assess every behavior or performance) were analytic scoring rubrics (Table 2). For the contents, most were general scoring rubrics, therefore this study took the same checklist and criteria in every scientific activity and a few targeted criteria were listed additionally. (For instance, to estimate the children could make more products or not in children's fluency assessing, and what amount of products is plenty depended on different scientific activity.)

Table 2

Children's scientific creativity scoring rubrics -- example "Shooting Arrow"

Item	Score	Criterion	Children's possible performance (instances)
Try different combinations in the process of performance assessment.	Goal：To understand children's multi-combinations.		
	2	Used more than 5 combinations.	Combined 2 straws: Combined 3 straws. Combined many straws.
	1	Used 2-4 combinations.	
	0	Used only one combination.	

2. Preschool Children's Science Flow Experience Scale

Children's curiosity and motivation was key factor of creativity. Therefore, the researcher developed a Preschool Children's Science Flow Experience Scale based on flow experience of Csikszentmihalyi (1996, 2000) in three stages.

The Children's scientific flow rating scale was a 3-point Likert scale, including 13 questions, separated in 5, 5, and 3 for three steps. After judging the degree of children's each behavior or performance, teacher gave 0, 1 or 2. This scale was used in every teaching unit.

3. Children's Scientific Knowledge Work Sheet

Before every scientific activity, researcher drafted a concept map by all variables and designed questions by every variable in the concept map. Every question was a scientific challenging question, and was also one activity. To avoid language interference in respondents, all questions appeared in pictures and were explained orally by teacher. Children answered the questions under Challenging Games with no time limitations. In the process, children had to make variable assumptions, do hands-on by themselves and record the result by colored pen. In scoring, score dichotomies were taken; one right answer could gain one point and zero for the wrong answer.

4. Creativity Products Evaluation Criteria

Creativity had its own purpose; the utility assessment of the product was an important element. Therefore, real competition took place in the last class of every teaching activity to let children apply their hands-on products, to check the function and to give self-feedback to individual problem solving skills. This product utility of every teaching unit had a different goal, they were: the farthest shooting arrow, the farthest air gun, the highest paper cup propeller, the steadiest wire-walker and the best tumbling jumping bean. And the scoring methods of every hands-on activity were different, for instance, in the competition of shooting arrow, it scored by scales which were marked by nylon ropes in the outdoor field. Each scale was one meter; the farther straw got a higher score. The competition was implemented five times, and the total was the score of product utility.

In Hou(2009), the above-mentioned scales had satisfactory construct-related validity evidences by using confirmatory factory analysis. In the pilot study, the utility and practicality were assessed by teachers, investigators and raters, questions were deleted and remained under their agreement. The interviews with teachers, parents and children conformed social validity of the course and assessment.

Analysis Methods

The present study adopted GEMOVA software (Brennan, 1992) and SPSS 17.0 to have data analysis of the generalizability theory and the correlation analysis. We would specify in the following.

1. From generalizability study, we would understand the generalizability coefficients (ρ^2) in performance assessment of children's creativity and the variance of dimension and task by p×d×t cross-over design.
2. From decision study, the generalizability coefficients (ρ^2) was above .80. We could thus make amendment to the content of the performance assessment.

Research Results

The task of performance assessment in this study were 5 hands-on physical activities, each task included four different scoring of response format, therefore researcher analyzed all the data in this performance assessment to estimate generalizability coefficient in 'task' and 'dimension' measurement facet in order to understand variance components and to have evidence in reliability. Interpretation of the result estimated by generalizability theory in pilot study and formal study was as below.

Tables 3 and Table 4 were the results of pilot study data estimation by the generalizability theory. Table 3 was three-way ANOVA and variance proportion estimation. There was only one datum in one cell, so it could not separate PTD. Furthermore, the biggest source of variable was dimension (σ_D^2 =28.731, 51%). The second variance proportion was from person (σ_P^2 =14.186, 25%). And, random errors were also noticeable (σ_E^2 =6.494, 11.45%). Other variance proportions were not large. Table 4 was generalizability coefficients and PHI coefficients of NT and ND. Due to the variances of different 'tasks' being small, giving more tasks was not helpful for raising generalizability coefficients, therefore using five tasks was appropriate. For raising generalizability, 'dimensions' the largest variable factor had to be enlarged, but according to the table, while T was 5, D would be 4, (ρ^2)= .890 was well for generalizability.

Due to large variation found in dimensions, it was almost more than twice that of person variability, researcher rechecked the scales and found that the score of Children's Scientific Knowledge Work Sheet were between 0 and 8. Comparing to other scales, the coefficient of variation was small and children got confused easily while answering the questions, for instance, children could express in words that jumping beans tumbled faster on the slope than on the plane that was also confirmed by their hands-on outcome, but there was not the same answer on the work sheet. Besides, while T=5 and D=3, there was .866 generalizability coefficients, more than 0.8, researcher decided to take this work sheet as a form of challenging question, not to count it into the score of scientific creativity.

Table 3

Three-way ANOVA and variance components estimation

Factor	df	SS	MS	F	Variance Components	Ratio
P	9	2868.21	318.69	49.08	14.19	25%
D	3	4439.18	1479.73	227.88	28.73	50.64%
T	4	63.78	15.95	2.46	0*	0%
PD	27	733.38	27.16	4.18	4.13	7.29%
PT	36	514.62	14.29	2.20	1.95	3.44%
DT	12	269.90	22.49	3.46	1.60	2.82%
error					6.49	11.45%

Comments : P is person（P=10） D is dimension（D=4） T is task（T=5）

*Actual value is -. 359, regarded as 0.

Table 4

Generalizability coefficients and PHI coefficients

Sample Size			G coefficients	PHI coefficients
P	D	T		
10	1	5	.709	.290
10	2	5	.820	.447
10	3	5	.866	.546
10	4	5	.890	.613
10	4	4	.880	.609
10	4	3	.864	.601
10	4	2	.834	.586
10	4	1	.755	.535

Table 5 was the generalizability theory estimation result of formal study. By variance proportion and ratio, dimension variable σ_D^2=18.75 taken 26.4%, person variable σ_P^2=15.03 taken 21.2%, random errors σ_E^2=18.00 taken 25.3% were the big source of variable. As to task × dimension which took 19.4% (σ_{DT}^2=13.79) meant different good performance in different task which scored in different dimension. In the part of generalizability theory estimation, while T=5, D=3, generalizability coefficients .811 was more than 0.8. The generalizability reliability in this performance assessment was confirmed.

Table 5
Generalizability theory estimation

Factor	**df**	**SS**	**MS**	**F**	**Variance Components**	**Ratio**
P	21	278.04	318.69	49.08	15.03	21.2%
D	2	2408.28	278.04	227.88	18.75	26.4%
T	4	147.60	2408.28	2.46	0*	0%
PD	42	42.62	147.60	4.18	4.92	6.9%
PT	84	27.98	42.62	2.20	3.32	4.7%
DT	8	321.51	27.98	3.46	13.79	19.4%
error	3025.248	168	18.01		18.00	25.3%
R Squared			.856	(Adjusted R Squared = .718)		
generalizability coefficients				.811		

Comments : P is person（P=22） D is dimension（D=3） T is task（T=5）
*Actual value is -2.79, regarded as 0.

According to the analyses above, different performance tasks were designed by the same concept of physical mechanics, therefore the task variable was almost zero, which confirmed the expectations of the researcher. Different dimensions had a big variable, because assessing dimensions included scientific creativity characteristics, flow and utility of product, these were different dimensions in theory construct, so the researcher estimated children's performances in different dimensions of scientific creativity by this multi-dimensions assessment. The variable of task × dimensions was big, which meant that different tasks emphasized different dimensions, the design of task and dimension could strictly embrace all possible creative performances of children. In the end, we still had to pay attention to the high error variance.

Conclusion and Suggestions

The performance assessment included performance task, response format, and scoring. The five science tasks in the performance assessment were easy activities to accomplish physically, and could attract children's curiosity and interest. They were helpful for the teachers to observe children's creative characteristics and science flow performance and to score the product. It's strictly concerned the purpose of creativity and obtained abundant formations of process and result. In considering the learning process and result in assessment, understanding of children's creative characteristics, flow and performance. In P×D×T cross-over design of generalizability coefficient estimation, variance components of dimension(D) and variance components of D×T were big, which meant that in different tasks one person's performance degree was different in these three dimensions, it showed the strict design of task and dimension embraced all possible creative performances of children. The whole generalizability coefficient was .811 with good generalizability reliability. It could be seen that taking five task activities and three dimensions were ideal combination. Thus the generalizability reliability in the performance assessment was confirmed. Significantly，the sample number only had 22, it may affect the outcome of generalizability.

Compared with the previous children's creativity assessment, the present study put domain and systems approach into consideration. Amabile (1995) and Csikszentmihalyi (1996) proposed that creativity should base on knowledge of specific domain. Runco (2006) also emphasized the concept of domain and it was similar to Craft's (2002) concept of "Little C". Gruber and Wallace (quoted from Ye, Lee, Ye, Lin & Peng, 2006) stressed that creativity was the interaction of individual, process and knowledge. Therefore, the present study not only gained insights into the characteristics of creativity, the motivation of the process, but also the utility of the product.

The scientific challenging questions in the knowledge sheet or orally expressed questions could promote children to increase their science creativity. Ye (2006) used context-based questions to assess children's creativity and found that four to six years old children had the ability of logical thinking and value judgments. This ability will increase with ages and promote through learning. While children immersed in the delight of hands-on activities, the teachers could make use of questions to let children thinking and try. To encourage the children had further thinking and manipulation, and the meta-thinking aroused by self-feedback would help them to internalize the correct science knowledge and develop science creativity.

The generalizability study provided children's science creativity performance assessment good reliability evidence, and it was worthy of further use.

References

Amabile, T. M. (1995). *Creativity in context*. CO: Westview Press.

Brennan, R. L. (1992). *Manual for GENOVA: A generalized analysis of variance system* [Computer software and manual]. University of Iowa

Brennan, R. L. (2001). *Generalizability theory.* New York: Springer-Verlag.

Brennan, R. L. (2011). Generalizability theory and classical test theory. *Applied Measurement in Education, 24*, 1-21.

Brooke, H. & Solomon, J. (1998). From playing to investigating: Research in an interactive science centre for primary pupils. *International Journal of Science Education*, *20*(8), 959-971.

Brown, J. H., & Shavelson, R. J. (1996). *Assessing hands on science.* Thousand Oaks: Cacorwin.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. NY: Harper Col.

Csikszentmihalyi, M. (2000). *Beyond boredom and anxiety: Experiencing flow in work and play* (2nd ed.). San Francisco: Jossey Bass.

Craft, A. (2002). *Creativity and early years education: A lifewide foundation.* NY: Continuum.

Doran, R., Boorman, J., Chan, F., & Hejaily, N. (1993). Alternative assessment of high school laboratory skills. *Journal of Research in Science Teaching*, *30*, 1121-1131.

Dunbar, S. B., Koretz, D., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4*(4), 289-304.

Feasey, R. (2005). *Creative science: Achieving the WOW factor with 5-11 year olds*. Landon: David Fulton.

Gardner, H. (1993). *Creating minds.* NY: Basic Books.

Guilford, J. P. (1950). Creativity. *American Psychologist*, 5, 444-454.

Hou, Ya-Ling. (2009). The technique development of assessing preschool children's creativity in science: An embedded assessment design. *Journal of Research in Education Sciences*, *54*(1), 113-142.

Kane, M. (2002). Inferences about variance components and reliability-generalizability coefficients in the absence of random sampling. *Journal of Educational Measurement, 39*(2), 165-181.

Kamii C., & DeVries, R. (1978). *Physical knowledge in preschool education: Implications of Piaget's theory*. EUA: Prentice-Hall International.

Linn, R., & Gronlund, N. E. (2000). *Measurement and assessment in teaching* (8th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Athins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal*, *38*(1), 73-95.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp.13-103). New York: Macmillan.

National Advisory Committee on Creative and Cultural Education (1999). *All our futures: Creativity, culture and education.* London: DfES.

Nickerson, R. S. (1999). Enhancing creativity. In R. J. Sternberg (Eds.), *Handbook of creativity* (pp. 392-430). Cambridge: Cambridge University Press.

Runco, M. A. (2006). The Development of children's creativity. In B. Spodek & O. N. Saracho (Eds.), *Handbook of research on the education of young children* (pp. 121-131). NJ: Mahwah.

Russ, S. W.（1994）. *Affect and creativity: The role of affect and play in the creative process.* New Jersey: Lawrence Erlbaum Associates.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer.* Newbury Park, CA: Sage.

Simonton, D. K. (1984). *Genius, creativity, and leadership*. Cambridge, MA: Harvard University Press.

Smoluch, F. (1992). The relevance of Vygotsky's theory of creative imagination for contemporary research on play. *Creativity Research Journal, 5*, 69-76.

Sternberg. R.J., & Lubart, T. I. (1995). *Defying the crowd: Cultivating creativity in a culture of conformity.* New York: The Free Press.

Sternberg, R. J. (1999). *Handbook of creativity*. New York: Cambridge University Press.

Stiggins, R. J. (1994). *Student-centered classroom assessment.* New York: Macmillan Publishing Company.

Wiggins, G. P.(1993). A*ssessing student performance: Exploring the purpose and limits of testing.* San Francisco, CA: Jossey-Bass Publishers.

Wiggins, G. P.(1998). *Educative assessment: Designing assessment to inform and improve student performance.* San Francisco, CcA: Jossey-Bass Publishers.

Wu, Jing-Ji.（吳靜吉）(2002). Chuangzaoli de yanjiu quxiang zh huigu yu zhanwang (The retrospect and prospect of creativity's research orientation). Paper presented at the Creativity Curriculum Development

Conference, Taipei: National Taipei University of Education.

Ye, Yu-Zhu, Li, Mei-Ling, Ye, Yu-Huan, Lin, Zhi-Zhe & Peng, Yue-Yin.（葉玉珠、李梅齡、葉玉環、林志哲、彭月茵）(2006). Youer chuangzaoli ceyan zh fazhan (The development of children's creativity). *Psychological Testing*, *53*(1), 129-154.