Development and Validation of an Instrument Assessing Preschool Children's Attitude Towards Science

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAII AT MANOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION
IN
EDUCATIONAL PSYCHOLOGY

August 2014

By
Tingting Reid

Thesis Committee:

Seongah Im, Chairperson

Leah Muccio

Lois Yamauchi

Abstract

This study intended to modify an established attitude instrument (ATSSA) to gauge preschool children's science attitude, with a focus on its psychometric characteristics (i.e., reliability and validity). The modified ATSSA scale consisted of 15 items that aim to capture 53 preschool children's interest in science and science related activities were read aloud. Children's responses were rated on a four-point scale ranging from 0(dislike) to 3(like it a lot). Results provided strong evidence for the reliability of the measure for both occasions (Cronbach's Alpha = .851and .822) and a unifactorial structure of the modified ATSSA scale. Further, confirmatory factor analysis demonstrated high loadings of the items onto the General Science Attitude factor, ranging from .453 to .859. Finally, generalizability coefficient (. 733) was obtained to demonstrate the moderately high degree of dependability of the current measurement design. Twenty items and three occasions were required to achieve the desirable reliability of .80. Implications and limitations of the study were also discussed.

Acknowledgements

Foremost, I would like to express my deep gratitude to my advisor Prof. Seongah Im for the continuous support of my Master program study and research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could never have imagined having a better advisor and mentor for my graduate study.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Lois Yamauchi and Prof. Leah Muccio, for their encouragement, insightful comments and hard questions.

My sincere thank also goes to my fellow quantitative study mate, James Tamayose, who has been immensely generous with his time to offer feedback on my thesis. I thank him for the stimulating discussions that inspired me to keep working hard.

Table of Contents

List of Tables

Children's development of attitude towards science begins early (Bruce C., Bruce P., Conrad, & Huang, 1997). Yet, little is known in regard to how young children feel about science in the preschool years (Mantzicopoulos, Patrick, & Samaraoungavan, 2008). It is of importance to scrutinize how they perceive their experience with science considering its influence on future behaviors, such as engagement in science-related activities, enrollment in advanced science courses, and choice of a career in science (Bruce et al., 1997). Osborne, Simon, and Collins (2003) characterized students' attitudes towards studying science as an "urgent agenda for research." The importance is highlighted by mounting evidence of a decline in the interest of young people pursuing scientific careers (Department of Education, 1994; Smithers & Robinson, 1988). Combined with the research indicating widespread scientific ignorance in the populace (Durant & Bauer, 1997), and an increasing recognition of the importance and economic utility of the scientific knowledge, the falling numbers of pursuing the study of science has become a matter of considerable social concern (e.g., Jenkins, 1994). Extensive efforts have been exerted to improve science education with emphasis only in middle school and high school; however, elementary school and Pre-K science have largely been ignored (Metz, 2011). It is upon this calling that I carried out a study assessing preschool children's attitude towards science and science learning.

Research has shown that a foundation for educational opportunities in science can help children promote children's learning in the subject (Saracho & Spodek, 2008). During the preschool years, children begin to construct science concepts of more complexity (Lind, 1998). Critical spans for knowledge attainment in young children occur between the ages of four and six, and for some subjects this critical window closes early (Begley & Hager, 1996). According to Eshach and Fried (2005), science is one of those subjects. In addition, the exclusion of preschool curriculum to literacy and language skills has been criticized in recent years (Mantzicopoulous et al., 2008). Because science receives substantially less attention, young children may not be familiar with what science is and thus limited evidence exists on their views about science as a content area.

One of the most important reasons for including science in early childhood education is because science taps into children's natural curiosity (Eshach, 2006; Esach & Fried, 2005; French, 2004; Rillero, 2005; Worth & Grollman, 2003). Children are more capable of reasoning in scientific ways than previously thought (Eshach & Fried, 2005), since for many years, science educators have considered preschoolers as irrational, illogical, "pre-causal," and limited to here and now based on Jean Piaget's developmental stages, the great pioneer of cognitive development (Piaget, 1929). However, recent theoretical and empirical research has begun to show otherwise; in that children's learning mechanisms resemble the basic inductive process of science (Gopnik, 2012). In this view, children are considered as "sophisticated scientists" (Metz, 2011) who possess extensive skills and capabilities in scientific hypothesis and reasoning. In addition, they are able to create theories about the world and how it works (Conezio & French, 2002). Such ways of thinking and skills are important to learning throughout life (Worth & Grollman, 2003).

Attitudes, such as value or liking of science, are associated with current and future science achievement and continued interest in science classes and careers (Jacobs, Finken, Griffin, & Wright, 1998; Simpson & Oliver, 1990). However, these findings are based on research with older children and adolescents and little is known about young children's attitudes toward science and science learning, as noted earlier. Further, this critical research is hampered by the lack of developmentally appropriate and psychologically robust instruments for measuring young children's attitudes in this regard (Mantzicopoulos et al., 2008). Therefore, one of the major objectives of this study is to address the gap in the early childhood science education by devising an attitude scale that can generate reliable and valid information regarding young children's attitude towards science and science learning.

Using the Eric, EBSCO, and Google Scholar databases, I searched the literature for combinations of the following keywords: preschool-age children, young children, science, scientist, attitude, images, ideas, nature, inquiry-based learning, cognitive capability, assessment, and instrument. In order to gain a better understanding of the challenges and problems in attitude research of science education as a whole, I looked through the bibliographies of the studies that predominantly involve the discussion concerning the construction of instruments. Hence, methodological issues emerged in quantitative studies regarding the measurement of attitude, such as definition of attitude, and psychometric properties of instruments, constitute the focal points of this review. In addition, qualitative methods, particular with a projective test, of assessing children's attitude towards science was also examined to some extent. Lastly, the review presented a synthesis of conclusions in terms of young children's attitudes towards science and science learning.

Literature Review

Definition of Attitude

As Osborne et al., (2003) claimed, one of the prominent challenges within the domain of attitude research in science education is a lack of clarity about the concept under investigation. If a scale is to be valid and reliable, there should be a preliminary attempt to specify as clearly as possible the theoretical construct underlying the scale. Attitude is such a broad term that is open to a number of interpretations. An early notable contribution towards its elaboration was made by Klopfer (1971), who categorized a set of affective behaviors in science education as:

- the manifestation of favorable attitudes towards science and scientists;
- the acceptance of scientific inquiry as a way of thinking;
- the adoption of ‘scientific attitudes’;
- the enjoyment of science learning experiences;
- the development of interests in science and science-related activities; and
- the development of an interest in pursuing a career in science or science related work

Further clarity emerged with the drawing of a fundamental and basic distinction by Gardner (1975). He distinguished two broad categories of attitude. The first category, “attitude towards science” (e.g., interest in science, attitudes toward scientists, attitudes toward social responsibility in science) shows some distinct science objects such as science, or scientists, to which the respondent is invited to react favorably or otherwise, which is the construct prevalent in those studies. The second category, “scientific attitudes” (e.g., open-mindedness, objectivity, honesty, and skepticism), by contrast, are best described as modes of thinking which scientists are presumed to display. Ormerod and Duckworth (1975) concur with such a distinction when they maintain that

> It is important to distinguish between those studies which investigate what can be described as a general attitude or disposition of mind for or against scientists and scientific activity, and other studies which are devoted to the identification or assessment of those desirable scientific attitudes—regard for evidence, thoroughness, attention to detail…” (Ormerod & Duckworth, p. 6).

The first of Gardner’s (1975) two categories concentrates on the emotional reaction that children might be expected to show towards science. It is on these emotional responses rather than more intellectual aspects developed through the study of science that are of interest. In this respect, Gardner regards attitude to science as "learned dispositions to evaluate in certain way objects, actions, situations or propositions involved in the learning of science" (p. 2).

Measurement of Attitude Towards Science and Science Learning

The following section draws on a range of attitude studies to discuss issues of how attitudes are measured, and the problems existing in attitude instruments. As noted above, a substantive amount of research on science attitude has been conducted with older children in formal settings, so when general references are made to studies of "attitude towards science," they focus on attitudes that are a product of students' experience with science. Recognition of the difficulty of measuring attitudes toward science comes in the diversity of methods researchers have taken in its measurement, which I will review beneath.

Most commonly, attitude towards science and science learning has been measured through the use of questionnaires that consist of Likert-scale items where children are asked to respond to statements on the questionnaire or survey (e.g., science is fun; I would enjoy being a scientist). Each item is a component in a summated rating scale that consists of a number of opinion statements reflecting either a favorable or unfavorable attitude to the construct being studied. The subjects are then normally offered a five-point choice consisting of "strongly agree/agree/not sure/disagree/strongly agree" or a three-point choice consisting similarly of "agree/not sure/disagree" to indicate their own feelings. The items are normally checked by a panel of experts in the field for content validity (e.g., Murphy & Beggs, 2003). Such items are normally derived from the free response answers generated by children, which is the major justification for their validity. These open responses are then reduced to a set of useful and reliable items that have been piloted and further refined by statistical analysis to remove those that fail to discriminate. Such scales have been widely used and extensively trialed, and are the major feature of research in this domain.

As Gardner (1995) claimed, many instruments suffer from significant problems as a good instrument needs to be internally consistent and unidimensional. Internal consistency is commonly determined through the use of measure commonly known as Cronbach's alpha and is often cited in much of the research literature on the measurement of attitude (Brennan, 1992). However, while unidimensional scales will be internally consistent (since they all measure the same construct), it does not follow that internally consistent scales will be unidimensional. That is because a scale may be composed of several clusters of items each measuring a distinct factor. In this situation, as long as every item correlates with some other items, a high Cronbach alpha will be obtained. Dimensionality of scales is usually tested using an appropriate statistical technique such as factor analysis. If a scale does measure what is purported to measure, then the variance should be explained by a loading on a unitary factor.

An attitude scale developed by Germann (1988), Attitude towards Science in School Assessment (hence referred to as ATSSA, see Appendix A), is one claimed to have robust reliability (Cronbach's alpha = 0.93). It has been adopted by a number of science education researchers (e.g., Ong & Ruthven, 2009; Peleg & Baram-Tsabari, 2011; Silver & Rushton, 2008). The construct of attitude is consistent with Gardner's (1975) definition, "the affect for or against a psychological object." It attempts to measure a single dimension of general attitude towards science.

More specifically, it gauges the degree to which students like or enjoy science as a subject in school; not any specific science courses (i.e., lecture, classroom, group work, homework or fieldtrip) that occur within science classes. In this perspective, attitude does not include scientific attitudes, attitudes towards scientists, towards methods of teaching science, or towards scientific interest. Nor does it include judgments of personal ability in science, the value of science to the individual, or the value of science to society.

However, this attitude scale was designed for older children who were in grade 7 and grade 8 and seldom been used among younger children. One study that adapted ATSSA using a four-point Likert scale to gauge children's attitude towards science and science learning of a younger age, in grade 1 and grade 2, has nevertheless failed to report the internal consistency data (Peleg & Baram-Tsabari, 2011). Similarly, the omission of reliability coefficient in Silver and Rushton's (2008) study concerning grade 5 children's attitude to science learning severely undermines the validity of statistical conclusions. According to Schibeci (1983) and Munby (1983), one major problem is that attitude instruments are of low psychometric quality, and most measures do not provide appropriate psychometric evidence of reliability and validity.

Another salient issue facing attitude research lies in the confusion of various theoretical constructs. As noted above, specifying the construct of object (attitude) is of great importance in constructing an instrument. Instruments have frequently been constructed which contain two or more logically and psychologically distinct variables; however, the distinctions are either not perceived, or ignored, and all the item responses are summed to yield a single score (Osborne et al., 2003). For example, an instrument entitled Attitude to Science Questionnaire (Murphy & Beggs, 2003), some of the items reflect attitude to science lessons and science-related activities ('I do not think we do enough science practical work in school', 'We do too much written work in science'), while others reflect respondents' views about their own capability in doing science and perceptions on the importance of science ('I find it difficult to understand science results', 'Science is an important part of everyday life). Including items such as those makes unexplored assumptions about the relationship among attitude, capability, and significance of science learning. Analysis ought to be enough to show that these are distinct variables, which ought to be measured separately and not mixed together into a single score.

The problem of interpreting the significance of these multiple components of attitudes towards science has been clearly identified by Gardner. He noted that in order for the scale a yield a meaningful score, the scale ought to reflect "respondents' position on some well-defined continuum" (Gardner, 1975, p.12). In short, if there is no single construct underlying a given scale, then there is no purpose served by applying a summated rating technique to produce a unitary score.

From this perspective, three different scales should be used to measure each construct of interest, and each gives differing degrees of emphasis. A prime example is the instrument developed by Mantzicopoulos et al., (2008) among kindergarten children in regard to their attitude towards science and science learning. Rather than measuring the singular construct, "favorable or unfavorable response to an object" (Gardner, 1975), the instrument consists of two other subscales measuring children's content knowledge in science (i.e., Specific Science Knowledge and Skills), and their perception of capability in science learning (i.e., Ease of Science Learning). In a similar vein, factor analysis was adopted as the primary means to verify the multi-dimensionality of the instrument pertaining to those three disparate variables.

Statistical Procedures Adopted in Validation of Attitude Measures

Investigations into attitudes about science have employed a wide variety of standard statistical procedures. Factor analysis is one of the most commonly used procedures in the development and evaluation of psychological measures. It is particularly useful with multi-item inventories designed to measure personality, psychopathology, attitudes, behavioral styles, cognitive schema, and other multifaceted constructs of interest to psychologists (Floyd & Widaman, 1995). Accordingly, the extant empirical research purporting to investigate young children's attitude uses factor analysis to identify the underlying dimensions of attitude, as assessed by a particular measuring instrument (e.g., Germann, 1988; Mantzicipoul et al., 2008; Kind, Jones & Barmby, 2007; Murphy & Beggs, 2003; Murphy et al., 2006).

However, as discussed earlier, a criticism highlighted in the literature (Gardener, 1975, 1995; Germann, 1988; Munby, 1983; Osborn et al., 2003) is the neglect of reporting psychometric properties of the attitude instruments. In addition, factor analysis is largely missing in several the studies in this review (e.g., Ong & Ruthven, 2009; Peleg & Baram-Tsabari, 2011; Siver & Rushton, 2008). For the ones that did, a lack of justification for choosing specific analytic techniques (i.e., four using principal component analysis and one with common factor analysis) used in the factor analysis has posed as one of the limitations of the study. As Floyd and Widaman (1995) noted, issues regarding the choice of particular analytic techniques used in a factor analysis appear to be the aspects of a factor analysis that are least understood by practicing scientists, and thus are often reported in insufficient detail.

Another salient issue concerning the use of factor analysis is data quality (Floyd & Widaman, 1995). The most direct way to ensure the quality of data is through careful item selection and item analyses (Haynes, Richard, & Kubany, 1995; Smith & McCarthy, 1995). Floyd and Widaman (1995) further posited that the psychometric properties of items to be factor-analyzed was of great concern, and pilot-testing of items should be performed to ensure that items designed to measure a common construct correlate moderately with one another and with the total scale score. For example, both Germann's (2003) and Mantzicopoulos (2008) carried out those procedures in terms of retaining the items in the questionnaire measures, with item-total correlations ranging from 0.13 to 0.83, and 0.61 and 0.89 respectively. However, the use of dichotomous response format (e.g., this child likes reading science books vs. this child does not like to read science books) in Mantzicopoulos's (2008) study might lead to biased results (Floyd & Widaman, 1995), and adequate analytic solutions to these problems of bias were not discussed.

Validity is also a key aspect when assessing the psychometric properties of an instrument. Failure to properly address construct validity is also a threat to good psychometric quality and there is a danger of ignoring validity in light of support from high consistency or reliability. A multitude of methods are adopted across the studies to examine the construct validity, such as the panel method (Germann, 1988), correlation coefficient to demonstrate convergent and divergent validity (i.e., theoretically similar items should converge and theoretically different items should be discriminating) (Kind, Jones, & Barmby, 2007; Mantzicopolous et al., 2008; Ong & Ruthvan, 2009;). Munby (1983) questions the assumption held in this technique that the meanings of items for the judges is the same as it is for the respondents. Similarly, Munby (1997) suggests using factor analysis to show that conceptually formed scales do in fact match with empirical produced factors (e.g., Mantzicopolous et al., 2008).

Qualitative Methods Employed in Assessment of Attitude

Attempts to measure attitude towards science and science learning have, in the main, shown a reliance on quantitative methods based on questionnaires. A common criticism of all attitude scales derived from such instruments is that, while they are useful in identifying the nature of the problem under investigation, they have been of little help in understanding it (Osborne, 2003), which has led to the growth of qualitative methods. Methods consisting of observation of in-class activities, field notes, video recording, surveys and semi-structured interviews are most prevalent in the qualitative studies to collect data (e.g., Bruce et al., 1997; Jalil et al., 2009; Sadi & Cakiroglu, 2011).

In addition, one prominent way suggested by a line of qualitative research in children's attitude towards science is through a projective test (drawing the images of scientists). Although scales of Likert types make no attempt to hide anything from the respondent, a deeper understanding can be gleaned through projective test since it is believed to "delve beneath the surface of superficial answers in an attempt to uncover hidden attitudes which are not revealed through ordinary methods" (Lowery, 1966, p. 35)

During the past few decades, there has been a growing body of research adopting such method to assess children's attitude towards science, because it is assumed that children's attitudes towards science are greatly influenced by their perceptions of science and scientists (Buldu, 2007). Those studies assessing perceptions children have of scientists by drawing (Chambers, 1983; Fort & Varney, 1989; Finson et al., 1995; Krause, 1977; Huber & Burton, 1995; Schibeci & Sorenson, 1983) have shown that children have a stereotypical image of scientists. They generally perceive scientists as males with glasses, beards and strange hair, and wearing white lab coats: individuals who work in laboratories messing with the chemicals or constantly scribbling notes in lab books. In these studies, the relationship between children's perceptions of science and scientists was confirmed to some extent; additionally, several children participating in the interview articulated their wishes of becoming the scientists they drew.

Even though drawing the scientist can be considered as a viable means to measure young children's attitudes towards science represented by their images of scientists in light of their limited reading and expressive skills, it is nonetheless not without drawbacks. Losh, Wike, and Pop (2008) pointed out that prior research adopting such method suffers from a major limitation, in that the neglect of making comparisons across occupations and failure in knowing how children view scientists as distinct from other professionals, interpolations to children's own academic or career motives are thus rendered suspect. Taking a slightly different approach, Losh et al., (2008) asked the children to draw images of three different professional practitioners and the results confirmed the previous findings that scientists do suffer an "image problem" that developed among young children, who consider scientists as largely white, and often unattractive men. However, the sensitivity of children's drawings (especially among the girls) to environmental elements and often-unrecognizable gender figures in boys' pictures raised skepticism about the projective personal inferences in relation to science identification from the prior research using such a methodology (Losh et al., 2008). Therefore, the projective test, such as drawing an image of a scientist, has been demonstrated to be a reliable measurement (Finson, 2003) yet warrants further evidence of validity.

Research Questions

In light of the previous discussion, the following guidelines on how best to formulate an attitude measure can be put forward: a) clear descriptions need to be put forward for the construct that one wishes to measure; b) reliability of the measure needs to be demonstrated by confirming the internal consistency of the construct (e.g., by use of Cronbach's alpha) and by confirming unidimensionality (e.g., by use of factor analysis); c) validity needs to be demonstrated by the use of more than one method, including the use of psychometric techniques.

This study modified the instrument, Attitude towards Science in School Assessment, developed by Germann (1988), and used it to assess children's attitudes towards science and science learning at four local preschool classrooms in Hawaii. In view of this background, questions of interest being addressed in the present study are:

1) How reliable is this modified scale of attitude towards science? In particular, what are the generalizability coefficients of the sub-scales?

2) How valid is this modified scale of attitude towards science? In particular, how high is total variance explained by attitude constructs? How high are factor loadings?

3) How positive are preschool-aged children's attitudes towards science?

Methods

Sample Characteristics

For the present study, I collected data from 71 preschool children from four Head Start preschool classrooms on O'ahu. Participants ranged in age from 36 months to 60 months, with a mean age of 53.76 (SD= 5.9 months) months. There were 36 females and 35 males. In addition, 48 participants were identified as English-only language learners. Ten participants were identified as dual language learners. Thirteen participants were not identified as either English or dual language learners. Slightly more than half the children (56%) were Native Hawaiian, 14% others were of other Pacific Island heritage, 26% were Asian American, and less than 2% were white. Fifty-six participants were exposed to a supplementary curriculum called World of Words (WOW). WOW is an enhanced and accommodated curriculum aimed at promoting economically disadvantaged young children's vocabulary and conceptual development. Fifteen participants were only exposed to the Head Start preschool curriculum.

Procedure

After receiving approval from the University's Institutional Review Board (IRB), consent forms were handed out to the parents of participants. The consent form notified parents that I planned to interview their child and included the questionnaire I intended to use during the interview. All parents of the participants returned the signed consent forms.

I scheduled two interviews that were two weeks apart through the lead teachers for the four preschool classrooms. After scheduling the interviews, I went into each classroom and selected one child to be interviewed. Before I conducted the formal interview, I had successfully established rapport with the children by spending a considerable amount of time in the classroom observing and interacting with them. For example, during meal times, I initiated conversations with these children about the topics they learned in the circle time that morning and participated in the play time.

Each interview took approximately twenty minutes. Before I started the formal interview, I performed validity checks by asking questions that related to the degree of interest. For example, I asked the child whether they liked pizza and the degree to which they liked pizza. If they answered “no,” I stopped there. If they answered “yes,” I continued by asking them the degree of their likeness. Further, I used hand gestures to verify children’s understanding of the degree of likeness. Once I was able to verify participants’ understanding, I asked the questions for the interview. After the individual child completed the questionnaire, I sent the child back to the normal classroom activity.

Instrument

The present study employed a modified version of the Attitude towards Science in School Assessment (ATSSA) developed by Germann (1983). The original ATSSA was used to gauge seventh and eighth grade children’s attitudes towards science and science learning (see Appendix A). The original ATSSA consisted of 14 items that are rated using a 5-point Likert Scale. The original ATSSA had demonstrated sound construct validity with item loadings greater than 0.40 for the construct of general attitude towards science (i.e., affect for or against science). Internal reliability was high with Cronbach’s alpha estimates greater than 0.95.

I modified the original ATTSA to accommodate the sample in my study. From the 14-item ATSSA, I changed the wording of the questions based on the WOW curriculum. For example, I modified the item, “I would like to learn more about science” to “I would like to learn more about insects.” I also added one item, “ I like using magnifying glass to observe insects or flowers.” Further, I reduced the number of response categories from five to four. In my modified ATSSA, participants were asked to rate items using a four-point scale, ranging from 0= dislike, 1= like a little bit, 2= like a little bit more, to 3= like a lot.

Data Analysis

Data were analyzed using SPSS 19 and Mplus 7.1.1 (Múthen & Múthen, 2013). I conducted the data analyses in a step-wise fashion, starting with missing data, descriptive analysis, internal consistency, Principal Component Analysis, Confirmatory Factor Analysis, and finally Generalizibility theory.

Missing Data. Listwise deletion was used to delete cases where complete data were not collected. Listwise deletion involves the removal of cases with missing values and excludes them from subsequent analyses (Enders, 2006). In my study, reasons for missing values include absences of participants, unexpected interruptions, and distractions. After the missing data were removed, the final sample analyzed consisted of 53 children. For the current sample, there were 27 females and 26 males. Ten children were dual language learners. Forty-one children were exposed to the WOW curriculum.

Item Reliability Analysis. Cronbach’s Alpha was used to check the reliability of my modified ATSSA scale. Values of Cronbach’s Alpha range from 0 to 1, with 0.7 considered acceptable reliability and 0.8 considered good reliability (Nunnally, 1978). After evaluating the scale reliability, individual items’ effect on scale reliability was conducted. More specifically, Cronbach's Alpha was computed with each item deleted to examine the item quality. Based on the Cronbach's Alpha estimates, items that weakened the reliability of the scale were removed.

Principal Component Analysis (PCA) and Confirmatory Factor Analysis (CFA). PCA was used to assess the dimensionality of the data of the modified ATSSA scale. Since eigenvalue is used to measure the amount of variation explained by each component, I selected the component that had the highest eigenvalue. Then, CFA was conducted to further examine and confirm the relationship between items and factor and the factor structure of the observed scores. Due to the categorical and ordinal nature of the items for my modified ATSSA scale, Weighted Least Square Mean Variance (WLSMV) was employed as the estimation method (Finney & DiStefano, 2006). To determine the fit of the 1-factor model, I focused on three fit indices, following the advice of Byrne (1998): the chi-square test, the Comparative Fit Index (Bentler, 1992), and the root mean square error of approximation (RMSEA) (Steiger, 1990). The non-significant result of chi-square test indicates a good fit. Additional fit indices were considered because the chi-square test is extremely sensitive to the sample size. CFI values > .90 indicate increasing good fit as they approach an upper bound of 1. The RMSEA reflects how close the model fit approximates reasonably fitted model, and indicates good model fit with values < .05.

Generalizability study (G study). A Generalizibility study was also conducted to determine the reliability of this scale, since it effectively identifies the multiple sources of measurement error. Moreover, G study is well suited to addressing such matters in that it enables the investigator to quantify and distinguish the sources of inconsistencies in observed scores that arise over replications of a measurement procedure. An $item \times person \times occasion$ $(13 \times 53 \times 2)$ crossed design was used in the present study, considering that item, and occasion are the two variables contributing to the sources of error. Although occasions were two weeks apart, no additional instruction or science related activities occurred between those two measurement occasions based on my observation. In another word, children's responses to the questions would not be affected by additional sources of variances, thereby reducing the error variances. G coefficients using both absolute error variances and relative error variances were calculated to provide a better understanding of the degree to which variability in the observed scores were due to measurement error. In essence, the construction of G coefficients was to make sure that the modified ATTSA scale was effective in measuring preschool children's general science attitude.

Results

Descriptive Statistics

Sample descriptive statistics such as the means and standard deviations for items for both measurement occasions are presented in Table 1. Higher mean values represent children's higher interests in science. Larger standard deviation values indicate a wider dispersion of interests in the subject matter of the item.

Table 1

Means and Standard Deviations Results for Modified Attitude Towards Science Questionnaire

	Time 1		Time 2	
Item	Mean	SD	Mean	SD
1.I like playing at the science table.	2.380	0.913	2.420	0.762
2.I like observing like a scientist.	2.040	1.018	2.180	1.090
3.I like using a magnifier to observe insects or flowers.	2.160	0.977	2.400	0.915
4.I am interested in reading books about marine mammals.	2.360	0.825	2.420	0.956

5.I would like to learn more about insects.	2.000	1.089	2.350	1.049
6.I would feel sad if I cannot play at the science table.	0.820	1.188	1.020	1.269
7.I enjoy reading books about insects.	2.160	0.958	2.090	1.093
8.I like to learn how to grow a plant.	2.250	1.022	2.560	0.788
9.I like learning about wild animals.	1.890	1.133	2.290	0.975
10.I like to ask questions about nature.	1.980	1.225	2.200	1.161
11.I like learning about what pets eat.	2.160	1.032	2.440	0.977
12.I like class discussions about pets.	2.110	1.083	2.240	1.088
13.I like observing the plants.	2.150	1.129	2.560	0.898
14.I like playing with dinosaurs.	2.050	1.096	2.420	1.031
15.I like talking about weather.	2.050	1.113	2.280	1.089
Average	2.040	1.050	2.260	1.010

Reliability Coefficients

Cronbach's Alpha was used to calculate the internal consistency of the modified *ATTSA* scale based on the covariation among items (Nunnally, 1978; Webb, Shavelson, & Haertel, 2006). For occasion 1, the Cronbach's Alpha is 0.851, and for occasion 2, the Cronbach's Alpha is 0.822. The function of the items as a test indicates a high level of reliability. Coefficients at or above 0.80 are often considered sufficiently reliable (Webb et al., 2006).

Estimates of Cronbach's Alpha when individual items are removed from the scale are presented in Table 2. For example, when item 1 "I like playing at the science table" is removed, the Cronbach's Alpha of the scale drops from 0.851 to 0.841. When item 6 "I would feel sad if I cannot play at the science table" is removed from the scale, the Cronbach's Alpha of the scale increases from 0.851 to 0.855. In general, items were found to have good quality, and internal consistencies for the whole test were found to be acceptable. This is positive evidence of consistency for the modified ATTSA scale.

Table 2

Summary of Reliability Coefficients

Item	Cronbach's Alpha if Item Deleted	
	Time 1	Time 2
1.I like playing at the science table.	.841	.825
2.I like observing like a scientist.	.836	.817
3.I like using a magnifier to observe insects or flowers.	.849	.812
4.I am interested in reading books about marine mammals.	.847	.825
5.I would like to learn more about insects.	.839	.801
6.I would feel sad if I cannot play at the science table.	.855	.839
7.I enjoy reading books about insects.	.845	.799
8.I like to learn how to grow a plant.	.850	.804
9.I like learning about wild animals.	.836	.802
10.I like to ask questions about nature.	.829	.806
11.I like learning about what pets eat.	.833	.814
12.I like class discussions about pets.	.835	.802
13.I like observing the plants.	.834	.800

14.I like playing with dinosaurs.	.845	.820
15.I like talking about weather.	.844	.803

Principal Component Analysis

Principal Component Analysis (PCA) results are presented in Table 3. PCA was used to identify the dimensionality of the scale with the current sample, because it is considered as an efficient way to sort manifested variables onto different dimensions and identify items that identify each factor (Widaman, 2010). In PCA, eigenvalues are generated from unreduced correlation matrix, characterized by 1's on the diagonal (Brown, 2006). The eigenvalues represent variance in the indicators explained by the component.

In Table 3, PCAs were conducted with 13 items, 14 items, and 15 items. For this modified ATTSA scale with 15 items, PCA generated four eigenvalues that were greater than 1. For the scale with 14 items where item 6 has been removed, PCA generated three eigenvalues that were greater than 1. Finally, PCA for the 13-item scale generated two eigenvalues greater than 1. I decided to use 13-item scale because the first component solely explained 45.848% of the total variance, and the ratio of factor one to factor two is about four, which confirms the unidimensional nature of the scale (Stout, 1990).

Table 3

Results of Principal Component Analysis with 13 items, 14 items, and 15 items

	15 items		Without item 6		Without item 3 and 6	
Component	Total	% of variance	Total	% of variance	Total	% of variance
1	6.212	41.414	6.107	43.623	5.96	45.848
2	1.752	11.682	1.656	11.828	1.579	12.148
3	1.438	9.590	1.435	10.253	0.952	

4	1.353	9.023

Confirmatory Factor Analysis

Based on the PCA results, I decided to use a modified ATSSA scale with 13 items where items 3 and 6 were removed. Confirmatory Factor Analysis (CFA) was conducted to further examine the internal structure of the scale in regards to item-factor relationships. Table 4 presents factor loadings and direct factor-item correlations. All items were meaningful indicators of the general science attitude factor, as indicated by factor loadings. Factor loadings ranged from .453 to .809, with 11 out of 13 direct factor-item correlations greater than 0.5.

Table 4

Summary of Item Loadings and Standard Error

Item	Factor Loading	Residual	Standard Error
1.I like playing at the science table.	0.591	0.651	0.109
2.I like observing like a scientist.	0.611	0.627	0.092
4.I am interested in reading books about marine mammals.	0.473	0.776	0.121
5.I would like to learn more about insects.	0.665	0.558	0.084
7.I enjoy reading books about insects.	0.567	0.679	0.096
8.I like to learn how to grow a plant.	0.453	0.795	0.129
9.I like learning about wild animals.	0.728	0.470	0.086
10.I like to ask questions about nature.	0.801	0.358	0.073
11.I like learning about what pets eat.	0.809	0.346	0.068
12.I like class discussions about pets.	0.779	0.393	0.072
13.I like observing the plants.	0.763	0.418	0.077

14.I like playing with dinosaurs.	0.534	0.715	0.107
15.I like talking about weather.	0.644	0.585	0.098

In assessing the model fit, the goodness-of-fit indices achieved adequate fit in the sample (χ^2=75.719, df= 65, p-value=0.1709 > .05, comparative fit index [CFI] =. 972, root mean square error of approximation [RMSEA] = .056, 90% confidence interval = .00 - .103). Although the point estimate of RMSEA .056 exceeds the threshold of .05, the value can still be interpreted as a good model fit (Brown & Cudeck, 1993; Hu & Bentler, 1999). In other words, 1-factor model adequately represented the relationship between the observed variables (i.e., 13 items) and the latent variable (i.e., construct of interest in science) in the present study. More specifically, the parameter estimates (i.e., factor loadings and residual variances) were reliable and warranted further investigation in regards to the quality of the items.

The sizes of the loadings reflect the extent of the relationship between each observed variables and factors (Tabachnick, Fidell, & Osterlind, 2001). The squared factor loading represents the proportion of variance in the indicator that is explained by general science attitude in the present study. Conversely, residual variances represent the proportion of variance in the indicators that is not explained by general science attitude. These errors (residual variances) were computed as 1 minus the squared factor loading. Therefore, high factor loading estimate reveals the indicator (item) is strongly related to the purported latent factor. In other words, items with higher factor loadings (lower residual variances) are more reliable.

In the present study, the three strongest indicators of the factor are items 10, 11 and 12, containing the least amount of error variances (See Table 4). Approximate 80% of variance in those three indicators respectively can be explained by the General Science Attitude factor.

Conversely, the two weakest indicators of the factor are items 4 and 8, containing the highest amount of error variances (close to 80%). More specifically, approximate 80% of the variance in those two items is measurement error, meaning only 20% of the variance can be explained by the General Science Attitude factor. Additionally, indicators that contain close to 50% of the error variances are items 9 and 13. Indicators that contain more than 50% of the error variances are items 1, 2, 5, 7, 14, and 15.

To elaborate, responses to items 10, 11, and 12 were greatest affected by the latent factor, children's general science attitude, since those three items tend to capture children's innate inquisitiveness of nature. However, responses to items 4 and 8 were severely affected by sources other than the underlying latent factor (i.e., measurement error). Hence, items 4 and 8 were highly unreliable. Ambiguities in the content of the item might have led participants to respond to different parts of the item. Another plausible cause for item unreliability could be differences in past experiences. For example, some children might have had some negative experiences associated with planting a tree and they were less likely to endorse a positive response accordingly.

In essence, aside from items 4 and 8 which contain around 80% of the error variance, all the other items are reliable in measuring children's general science attitude. Moreover, the 11 items demonstrated relatively high level of validity since they adequately measured what they were purported to measure. Therefore, a single-factor (i.e. general science attitude) general model was sufficient to account for the covariances in the observed scores.

Generalizibility Theory

Table 5 provides the estimates of the variance components. This design consisted of seven variance components: person, item, occasion, person by item, item by occasion, person by occasion, person by item by occasion. The person by item by occasion variance component refers to the residual variance. The residual variance is the variance that cannot be attributed to a single source of variation.

The variance associated with persons accounted for 10.77% of the total variability of item scores. Approximated 11 percent of total variability in children's science attitude scores was attributable to their innate systematic differences. Estimates of variance components for facets were .73% for items and 1.83% for occasions. The variability due to items and occasions respectively were negligible, meaning there is a high level of stability for both items and occasions. Variance estimates for interactions were 1.92%, 0.00%, and 18.12% for person by item, item by occasion, and person by occasion interactions respectively. Variance components associated with persons contributed to a sizable amount of variability in children's science attitude scores. For my study, explaining changes in children's responses across measurement occasions was difficult due to the confounded error variances. Additionally, the residual variance accounted for 66.61% of the total variability (Table 5). A large source of variability in children's science attitude scores was attributable to measurement error, meaning that identifying the single source of variability was difficult.

Large measurement error variances led to low Generalizibility coefficient (G coefficient) since G coefficient score reflect the ratio between the universe score variance and expected score variance (i.e., summation of universe score variance and residual variance). Analogous to the reliability coefficient from Classical Test Theory (Brennan, 1992), values for the G coefficient range from 0 to 1.0. Values that approach 1 indicate the universe score variance accounts for the main source of variability in the observed score variance. Conversely, values that approach 0 indicate that sources of variability other than universe score variance account for the variability of the observed score.

In my study, the person variance component represented the universe score variance (see Table 5). The observed score variance was the summation of the universe score variance and the remaining six variance components, also known as the absolute error variance. Using the absolute error variance, I was able to compute the generalizability coefficient (.733). More specifically, 73.3% of the variability of children's science attitude score was attributable to their systematic differences. Another aspect of generalizability theory is the calculation of generalizability coefficient regarding relative error variance. The relative error variance only concerns the variances associated with persons, which included person by item, person by occasion, as well as person by item by occasion interactions. I computed the generalizability coefficient using relative error variance (.799). This value suggests that approximately 80% of the variability of the observed children's science attitude score was attributable to systematic differences within the universe score.

Table 5

Estimates for G-study Variance Components of Two-Facet Crossed Design

Source of Variablity	df	MS	VC	%VC
persons(p)	52	6.431	0.118	10.77%
items(i)	12	1.582	0.008	0.73%
occasions(o)	1	17.21	0.020	1.83%
person* item (pi)	624	0.773	0.021	1.92%
item*occasion(io)	12	0.729	0.000	0.00%
person*occasion(po)	52	3.313	0.199	18.12%

person*item*occasion(pio)	624	0.73	0.73	66.61%

$\rho_{\Delta} = .458$ $\rho_{\delta} = .478$

One of the major advantages of generalizability theory is to permit decisions (called D studies) about the optimal types of designs used in assessing children's science attitude scores. The preliminary G study suggests that one major improvement in accurately assessing preschool children's general science attitude would result from increasing the number of occasions (i.e., person by occasion interaction contributes to 18.12 percent of the variability of the observed scores). Although the facet, item, only contributes 0.73 percent of the variability in the observed scores, I decided to include it in the D studies to explore the optimal combinations of facets that yield dependable information about children's general science attitude. As summarized in Table 6, the G coefficient (ρ^2_{Δ}) would reach as high as 0.705 if I increase the number of items to 19 and occasions to 5. Additionally, it would be more effective to increase the number of occasions than items to improve the G coefficient considering the magnitude of improvement when the items were held constant.

Table 6

Given a Random Effects G Study p ×i × o Design

$n_o =$	3	3	3	4	4	4	5	5	5
$n_i =$	13	16	19	13	16	19	13	16	19
$\sigma^2_{(\Delta)}=$	0.557	0.568	0.576	0.64	0.65	0.658	0.688	0.698	0.705

Discussion

This section opens with the summary of results along with their interpretations. This is followed by strengths, weaknesses, and limitations of the study. Further, implications and contributions to the field are discussed. Finally, this discussion section closes with recommendations and guidelines for future research regarding early childhood science education.

Summary of Results

This study set out to develop and validate a modified version of Attitude towards Science in School Assessment (ATSSA). For this study, the modified ATSSA was found to be highly reliable as demonstrated by Cronbach's Alpha coefficient of 0.851 and 0.841 for my two measurement occasions. For single measurement occasions, the high reliability coefficient suggests that the items for the modified ATSSA function consistently as a group. For my two measurement occasions, the modified ATSSA was able to generate similar Cronbach's Alpha coefficients.

After evaluating the Cronbach's Alpha coefficients, I also examined reliability at the item level and found that two of my items, item 3 and item 6, were undermining the reliability of the scale. I believe item 3, "I like using a magnifier to observe insects or flowers," was weakening the reliability because of the double-barreled nature of the item. Children's response to the question can be either triggered by the word "insects" or "flowers" depending on their prior experiences. In other words, I cannot pinpoint where their responses originate. The ambiguity of the item presents challenges to interpreting the consistency of the responses in relation to other items in the scale. Additionally, negative wording may have had an effect on reliability. Although my modified ATSSA scale only included one negatively worded item (item 6), the reliability analysis indicated that removal of item 6 would improve scale reliability. Further, research suggests that negatively worded items may affect the factor structure of the scales (Herche & Engelland, 1996; Kelloway& Barling, 1990), has been demonstrated to require more cognitive resources (Barnette, 2000; Merritt, 2012), may cause participants to respond carelessly and create confusion (Johnson, Bristol, & Schneider, 2011; Roszkowski & Soven, 2010).

The concerns arising from item 3 and item 6 prompted me to further investigate the dimensionality of the modified ATSSA scale. After running Principal Component Analysis, the effect of item 3 and item 6 on the factor structure of the scale was evident. Since participants could have responded to two different aspects of the content within item 3, "flowers" or 'insects"; this may have caused item 3 to load onto multiple components. When items have high cross loadings, interpreting the components becomes difficult. Item 6 also loaded onto multiple components, in a different way from item 3. The difference between item 3 and item 6 appears to be strictly due to the negative wording. Since interpreting the components was difficult, item 3 and item 6 were removed. With the removal of item 3 and item 6, I was able to identify a general science attitude component.

In determining the dimensionality of the component structure, I conducted Principal Component Analysis and found one predominant component. I hence named this component General Science Attitude. Consistent with Gardner's (1975) definition of "attitudes towards science," general science attitude refers to the natural inclination children displayed towards science and science related activities. For my scale, General Science Attitude was measured by children's responses towards reading, learning, discussing and observing. Looking at the scale, science attitude seems to be strongly expressed through inquiry, such as discussions and reading books about nature, pets, and plants. General science attitude seems to be moderately expressed through more active, hands-on participation related to science, such as emulating the behaviors of scientists. Inquiry about science appears to be stronger than reading about science in the preschool children's formation of science attitude. Evidence for this was supported by the stronger loading for item 5 (0.665), “I would like to learn more about insects,” when compared to item 7(0.667), “I enjoy reading books about insects,” and item 4 (0.473), “I am interested in reading books about learning mammals.”

Another purpose of this study was to investigate the generalizability of the modified ATSSA. Findings from the generalizability theory analysis suggest that a sizable amount of variance, approximately 21% of the estimated variance, came from person by facet (item or occasion) interactions. Among the interaction between item and occasions, the small value of 0.27 % indicated that the variance of items across occasions was negligible. However, the large value of residual variance (68.6%), which again represented the three-way interaction among person, item, and occasion confounded with unsystematic or unmeasured error, suggest that children’s responses cannot be attributed to a single source of variation. Due to the inconsistencies of children’s answers, the generalizability coefficient was low (0.449), meaning that this modified ATSSA scale does not permit me to generalize preschoolers’ responses to questions regarding the general science attitude.

Finally, with regards to preschool children’s general interest in science and science activities of my present study, they showed a high level of interest as indicated through the mean scores of both measurement occasions. More specifically, children tended to exhibit a higher level of interest during the second measurement occasion. A possible explanation to this outcome could be that children were more familiar with the questions and hence were likely to respond more positively.

Strengths, Weaknesses and Limitations

Since Classical Test Theory tended to treat every individual the same and hence difficult to interpret the error, I decided to apply Generalizability theory which provides an advantage over Classical Test Theory (CTT). Roughly translated, Generalizability theory allowed me to take into account individual differences inherent in children. In the context of the children's natural capacity to think like scientists, CTT would over-simplify children to one type of scientist; whereas Generalizability theory enabled me to perceive children as qualitatively different types of scientists.

Further, the employment of a crossed design was able to fully realize the power of the Generalizibility theory (Brennan, 1992), since I was able to maximize the number of design structures that can be considered for future studies. In my study, by using a two-facet crossed design, I was able to measure children's general science attitude on more than one condition (item and occasion). The two-facet crossed design estimated seven variance components; in other words, it allowed me to explicitly recognize the multiple sources of variance in measuring children's science attitudes. The result of using this specific design was to facilitate the interpretation of children's science attitude scores.

Sample size was a major weakness in my study. Although I collected data from 80% of children in each classroom, when evaluating the model fit, the mixed outcome reflected through the confidence interval (.000 and .103) of Root Mean Square Error of Approximation (RMSEA) indicated the need for a larger sample size in order to obtain more precise results. Browne and Cudeck (1992) noted that if the upper bound level of the 90% confidence interval exceeds a certain value (0.1), the model was considered as a "poor fit," and if the lower bound level of the 90% confidence interval approaches zero, the model was then considered as a "good fit". In my study, the upper bound level of the 90% confidence interval exceeds 0.1 and the lower bound level of the 90% confidence interval was zero. This seemingly contradictory result was probably subject to a fair amount of sampling error because it is just as consistent with the close-fit hypothesis as well as the poor-fit hypothesis. A larger sample size was hence needed to warrant the model more confidence.

Although the results from RMSEA suggest a need for a larger sample size, gathering more participants to increase model confidence was not feasible. Secondly, recruiting more participants would improve the stability of the estimates for the variance components and reliability coefficient. However, the addition of participants would require me to consider contextual differences such as classroom, teacher, and curriculum. For example, if I were to recruit participants from another site, I would have to assess the homogeneity of the original and newly added participants. If I overlook the homogeneity of the combined sample, I am susceptible to estimating unreliable variance components. Ultimately, the trustworthiness of the measurement design might be compromised.

Another weakness of the study was the use of a two facet crossed design. The residual variance, which was represented by the three-way interaction between person, item, and occasion, was the largest source of variability in the observed scores. In situations where the measurement error is substantial, parsing out multiple sources of error becomes difficult. In addressing the issue regarding substantial measurement error, I could increase the number of items or increase the number of occasions. In the event that these two solutions are unable to reduce the size of the measurement error, I would consider making changes to the design. Hence, to strengthen the design, changes that more precisely capture the variability of the observed scores will be made.

Capturing consistent results across time among preschool children is an inherent challenge in early childhood measurement studies. There are a multitude of variables that affect the consistency of preschoolers' responses, such as age (Sabbagh & Baldwin, 2001), culture

(Carlson, Moses, & Lee, 2006), and content delivery method of teachers when implementing word learning strategies (Roskos & Burstein, 2011). Even though it was difficult to measure the development trajectory, children nevertheless tend to exhibit a persistent response pattern. Considering the primary purpose of this study, which was to validate the scale, the measurement of growth pattern was beyond the scope of this investigation.

An additional weakness of the present study concerned the challenges of maintaining young children's interest during the interview due to their short attention span. Although modifications from the original ATSSA scale were made to accommodate preschoolers, the number of items may have had an effect on children's responses. Further, the content of the questions may have affected interests differently within the preschoolers. Pinpointing the source of preschooler's interest in science using the content of items was difficult; as Schiefele (1991) noted, interest can be construed not only as a process of learning, but also an outcome. For example, a response to the item "I like reading books about marine mammals" could be interpreted based on past experiences or future expectations.

A final concern involves the influence of children's over zealousness about science. In my experience of the study, the majority of children displayed the excitement through their tone of voice, body movements and facial expressions. Overzealousness might be conducive to the acquiescent response style ("yea-saying"). An acquiescent response style refers to the tendency of responding by selecting socially desirable categories regardless of item content or the context (White, Leichtman, & Ceci, 1997; Williams, Williams, & Beck, 1973). "Yea-saying" is a form of acquiescent response style in which participants respond affirmatively with "yes", "yeah", or "yay" (Vaerenbergh & Thomas, 2013). Further, young children tend to say "yes" when they understand the questions; whereas they tend to say "no" when they do not understand the questions (Frutzley & Lee, 2003). Despite of the efforts made to constrain acquiescent response style, such as the use of one negatively worded item that was later deleted and reassurance of safe response, it appears that children's overzealousness of science was still present to some extent.

Children participating in the present study were exposed to the World of Words (WOW) curriculum, which emphasized acquisition and categorization of vocabulary (Neuman, 2011). The items on the modified ATSSA questionnaire were adapted in accordance with the WOW curriculum. Although I conducted validity checks by asking participants whether they understood the content of the item, there was a possibility that the participants were unable to comprehend the meaning of the item. For example, when asking participants, "I like using the magnifying glass to observe insects or flowers", participants may know what a magnifying glass is but may not know the use of it.

Differences in curriculum across preschool settings limit the generalizability of these findings to settings where the WOW curriculum and perhaps settings where similar curriculums have been applied. More specifically, differences arise from the emphasis WOW places on vocabulary learning and acquisition in regards to science. WOW has been used as an intervention program focusing on promoting economically disadvantaged preschool children's vocabulary and conceptual development through richly structured taxonomic categories. Children in my present study who had been learning WOW curriculum would be more familiar with the words in the questionnaire and thus tend to display a higher level of enthusiasm when responding to the questions. In addition, Neuman, Newman, and Dwyer (2011) suggest that children who are exposed to WOW are more engaged and enthusiastic about inquiry. In my study, engagement and enthusiasm are manifested through the increased usage of vocabulary, such as observe, discuss, and weather. Further, other researchers who intend to conduct similar studies may not replicate my experience.

Contributions and Implications for Future Research

I have found that there is little research in the area of measurement regarding early childhood education. This study would be valuable in that it was able to address the gap by establishing a reliable and valid scale assessing preschool children's general attitude towards science. In addressing the gap, I selected items that I believed to precisely measure science attitude. Similar to the original ATSSA scale, this modified ATSSA scale was demonstrated to be also reliable and valid. More specifically, the 13 items on the modified ATSSA were found as strong indicators of the general science attitude as the 14 items on the original ATTSA scale. In other words, these 13 items constituted a scale that covered the breadth and depth of science attitude among preschool children as well as the original ATTSA that purported to assess 7th and 8th graders' science attitudes.

The reliability coefficients in the form of Cronbach's Alpha and construct validity as evidenced by PCA and CFA further supported the soundness of the modified ATTSA scale that was administered. I also investigated the scale further by applying the Generalizibility theory. The purpose of the Generalizibility theory was to disentangle the sources of error variance that validated the reliability of the items. In addition, the framework of Generalizibility theory allowed me to examine the efficacy of the measurement design (i.e., item by person by occasion) in regards to the extent to which I could replicate these preschool children's responses across contexts. In essence, this study shed light on the salient issues regarding the measurement of preschool children in general as evidenced by the dearth of literature in this field.

This study makes a second contribution to the field by supporting previous studies regarding children as “natural scientists” (Eshach & Fried, 2005; Gopnik, 2012; Metz, 2011). I found in my study that children’s responses to items measuring children’s innate sense of inquiry revealed that children were capable of thinking like a natural scientists. In addition, responses to items that reflect children’s capacity to emulate scientists indicated that children were cognizant of how scientists behaved. For example, item 3, “I like using the magnifying glasses to observe insects and flowers,” represents the stereotypical behaviors that scientists are most likely to adopt. Investigations into children’s innate capacity to be natural scientists needs to apply more thoroughly designed approaches.

Although the addition of occasions can be considered as a viable means to reduce the variability in children’s attitudes scores, yet it risks introducing extraneous variance, such as children’s maturing process. When designing future studies, the number of occasions needs to be considered. Further, in addition to the number of occasions, researchers should come up with a more structured interview protocol to facilitate children’s respond consistently to the items on the questionnaire. Finally, if we added children from non-WOW classrooms, researchers can gain insight into the efficacy of WOW has on children in terms of their attitudes towards science and science learning.

In order to improve the quality of measuring preschool children, I recommend the following approaches. First, regarding items, I advise against using negatively-worded items and items that may require a significant amount of cognitive load on children. The use of negatively-worded items may confuse children. Item that use the words “and” and “or” may require children to think more about the content of the items, which may affect responses. Second, regarding the design of the study, more observations are needed to grasp a better understanding of the sources of variation. The inclusion of more facets, such as raters or teachers' effect, may allow researchers to gain a clearer picture of how young children feel about science. Finally, I encourage using additional aspects of qualitative approach to gain deeper insight into preschool children's innate capacity to think like scientists. One such example would be the use of semi-structure interview that allows for children's practice of their scientific thinking.

In conclusion, measuring preschool children's attitudes toward science is inherently a daunting yet promising task. Children are natural scientists who have the ability to display such disposition through their manifested interest in inquiry, reading, and behaviors that emulate scientists. The modified ATSSA scale has been demonstrated to be reliable and valid. However, researchers still need to develop better ways to assess the fluctuation in children's responses. Future studies need to use more structured questions to improve the preciseness of measuring preschool children's attitudes toward science.

Appendix A. Synthesis of Attitudes Instruments in the Empirical Studies

Author	Participants	Instrumentation	Construct of Attitude towards science	Statistic Analytical Tool
Murphy & Beggs, 2003	1,000 children; 57% younger children (8/9 years old) and 43% in the younger age group (10/11 years old)	Questionnaire (3-point scale); attitude items adapted from a survey of attitudes towards ICT among primary school children (no reliability coefficient)	Enjoyment of science; Appreciation of the importance of science; perceived ability to do science	Exploratory factor analysis (using principal components analysis with Varimax rotation)
Kind, Jones, & Barmby, 2007	11-14 age students from five secondary schools (932 students for pre-measures, and 668 students for post-measures)	Questionnaire (5-point scale, Cronbach $\alpha > 0.7$); items from the Relevance of Science Education Questionnaire, the 2003 Programme for International Student Assessment (PISA) questionnaire, and from the attitude to science for 5-11 year olds developed by Pell & Javis (2001)	Learning science in school; self-concept in science; practical work in science; science outside of school; future participation in science; importance of science; general attitude towards science; combined interest in science	Principal Component Factor Analysis
Germann, 1988	Pilot testing with a group of 125 science students in grade 7 and 8	Questionnaire (5-point scale, Cronbach $\alpha > 0.95$) with 14 items; Attitude toward Science in School Assessment (ATSSA)	General attitude towards science	Principal Component Factor Analysis
Silver & Rushton, 2008	120 children in grade 5	Questionnaire (5-point scale with 'smiley-face', adapted from West, Hailes & Sammons 1997; Pell & Javis 2001; Javis & Pell 2002); no reliability coefficient; word description and drawing images of	General attitude towards science	Chi-square test was used to establish whether the differences in those distributions were statistically significant.

		scientists		
Ong & Ruthven, 2009	186 male and 201 female students from 2 smart schools and 184 male and 204 female from 2 mainstream schools in Malaysia (ages n/a)	Malay version of Attitudes Towards Science in School Assessment adapted from the instrument developed by Germann (1988) with 11 items (test-retest and Cronbach reliability were found to be at 0.93 and 0.90)	General attitude towards science	Analysis of Covariance (ANCOVA); independent t-test used to compare the ATSSA (M) scores of students; ATSSA (M) data screening
Mantzicopoulos, Patrick, & Samarapungavan, 2008	113 kindergarten children from two schools; 56.6% Caucasian; 12.4% African American; 23% Hispanic; 8% other; free-lunch students (71.2%)	Puppet Interview Scales of Competence in and Enjoyment of Science (PISCES, Cronbach $\alpha = .84$); the pool of 30 items (bipolar statement) was derived after 1) review of items in self-concept scales for young children (e.g., Chapman & Tunmer, 1995; Harter & Pike, 1984; Marsh et al., 1998, 2002); 2) classroom observations; Children's narratives about science learning (theme coding)	General Science Competence and Specific Science knowledge and skills (27%); Science Liking (9%); Ease of Science Learning (5%).	Chi-square tests used to examine the comparability of the children in two schools on gender, ethnicity, and free or reduced lunch status; Common factor analysis comparing the oblique solution against an orthogonal solution.
Murphy, Ambusaidi, & Beggs, 2006	944 primary school children of 9-12 years old in Oman; 979 children from Northern Island (age n/a); 50.2% female and 49.8% male students in both samples	Questionnaire (3-point scale, Cronbach $\alpha = .79$); the pool of 18 items were developed by Murphy and Beggs (2003); Teacher-student discussion (i.e., feeling towards science in and out of school, future participation in science in post-	Enjoyment of science; Appreciation of the importance of science; Perceived ability to do science	Exploratory factor analysis

		secondary schools)		
Peleg & Baram-Tsabari, 2011	118+99 students from 2nd and 4th grade classes in public school, and 288+148+164 students from 1st grade to 6th grade	Questionnaire (4-point scale regarding to general attitudes to science); Cronbach α n/a; semi-structured interview	General attitudes towards science and science learning (e.g., Is learning science difficult?)	Average score was calculated for each question item and t test was used to examine significance
Sadi & Cakiroglu, 2011	140 (71 girls, 69 boys) 6th-grade students who were 12 years attending four whole classes in a public elementary schools in Turkey.	Questionnaire consists of 15 items and rated on a 5-point Likert scale. Reliability of SAS (Science attitude Scale developed by Geban, Ertepinar, Yilmaz, Atlan and Sahpaz, 1994) was 0.82.	N/A	N/A
Jalil, Sbeih, Boujettif, & Barakat, 2009	271 grade level one (6 years old) to grade level four (10 years old) participated in the study.	Qualitative study (surveying and interviewing the students);	Mono definition of attitude (simple emotional disposition to like/dislike) towards learning science	N/A
Bruce C., Bruce P., Conrad, & Huang, 1997	48 children, ages 5-12	Qualitative study (surveying and interviewing the students). Field notes and observation	N/A	N/A

Appendix B. Attitude toward Science in School Assessment

Please use this scale to answer the following questions:

SA - Strongly Agree A- Agree N-Neither Agree or Disagree D- Disagree SD - Strongly disagree

(Circle one choice.)

(1)SA A N D SD Science is fun.

(2)SA A N D SD I do not like science and it bothers me to have to study it.

(3)SA A N D SD During science class, I usually am interested.

(4)SA A N D SD I would like to learn more about science.

(5)SA A N D SD If I knew I would never go to science class again, I would feel sad.

(6)SA A N D SD Science is interesting to me and I enjoy it.

(7)SA A N D SD Science makes me feel uncomfortable, restless, irritable, and irritable.

(8)SA A N D SD Science is fascinating and fun.

(9) SA A N D SD The feeling that I have towards science is a good feeling.

(10) SA A N D SD When I hear the word science, I have a feeling of dislike.

(11) SA A N D SD Science is a topic which I enjoy studying.

(11) SA A N D SD Science is a topic which I enjoy studying.

(12) SA A N D SD I feel at ease with science and I like it very much

(13) SA A N D SD I feel a definite positive reaction to science.

(14) SA A N D SD Science is boring.

Appendix C. Modified Attitude toward Science in School Assessment

Questionnaire

1) 0 1 2 3 I like playing at the science table.

2) 0 1 2 3 I like observing like a scientist.

3) 0 I 2 3 I like using a magnifier to observe insects or flowers.

4) 0 1 2 3 I am interested in reading books about marine mammals.

5) 0 1 2 3 I would like to learn more about insects.

6) 0 1 2 3 I would feel sad if I cannot play at the science table.

7) 0 1 2 3 I enjoy reading books about insects.

8) 0 1 2 3 I like to learn how to grow a plant.

9) 0 1 2 3 I like learning about wild animals.

10) 0 1 2 3 I like to ask questions about nature.

11) 0 1 2 3 I like to learn what pets eat.

12) 0 1 2 3 I like class discussions about pets.

13) 0 1 2 3 I like observing the plants.

14) 0 1 2 3 I like playing with dinosaurs.

15) 0 1 2 3 I like talking about weather.

References

Begley, S., & Hager, M. (1996). Your child's brain. *Newsweek*, 127(8), 54.

Buldu, M. (2007). Young children’s perceptions of scientists: a preliminary study. *Educational Research,* 48(1), 121-132.

Brennan, R. L. (1992). Generalizability theory. *Educational Measurement: Issues and Practice*, 11(4), 27-34.

Bruce, C. B., Bruce, P. S., Concrad, L. R., & Huang. H. (1997). University science students as curriculum planers, teachers, and role models in elementary school classrooms. Journal of Research in Science Teaching, 34(1), 69-88.

Conezio, K., & French, L. (2002). Science in the preschool classroom: Capitalizing on children's fascination with the everyday world to foster language and literacy development. Young Children, 12-18.

DEPARTMENT FOR EDCUATION (1994). Science and math: a consultation paper on the supply and demand of newly qualified young people (London: Department for Education).

Durant, J. & Bauer (1997). *Public Understanding of science: the 1996 survey*. Paper presented at a seminar at the Royal Society, 8 December 1997.

Enders, C.K (2006). Analyzing structural equation models with missing data. In G. R. Hancock & R.O. Mueller (Eds.), Structural Equation Modeling: A Second Course, 1st ed, (313-342). Greenwich, CT: Information Age Publishing.

Eshach, H., & Fried, M. (2005). Should science be taught in early childhood? *Journal of Science Education & Technology*, 14, 315-336.

Eschach, H. (2006). Science literacy in primary schools and pre-schools. Netherlands: Springer.

Finson, K. D. (2003). Applicability of the DAST-C to the images of scientists drawn by students of different racial groups. Journal of Elementary Science Education, 15(1), 15-26.

Finney, S. J. & Di Stefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R.O. Mueller (Eds.), *Structural Equation Modeling: A Second Course*, 1st ed, (269-312). Greenwich, CT: Information Age Publishing.

Floyed, F., & Widaman, K. (1995). Factor Analysis in the Development and Refinement of Clinical Assessment Instruments. *Psychological Assessment*, 7(3), 286-299.

French, L. (2004). Science as the center of a coherent, integrated early childhood curriculum. *Early Childhood Research Quarterly*, 19, 138.

Gardner, P. L. (1975). Attitudes to science: a review. *Studies in Science Education* 2: 1-41.

Gardner, P. L. (1995). Measuring attitudes to science. *Research in Science Education*, *25*, 283–289.

Germann, P. J. (1988). Development of the attitude toward science in school assessment and its use to investigate the relationship between science achievement and attitude toward science in school. *Journal of Research in Science Teaching*, *25*(8), 689–703.

Gopnik, A. (2012). Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications. *Science* 337, 1623.

Hancock, G. R., & Mueller, R. O. (Eds.). (2010). The reviewer's guide to quantitative methods in the social sciences. Routledge.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*,238-247.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: *A Multidisciplinary Journal*, 6(1), 1-55.

Jalil, A. P., Abu Sbeih, Z. M., Boujettif, M., & Barakat, R. (2009). Autonomy in Science Education: A Practical Approach in Attitude Shifting Towards Science Learning. *Journal of Science Education and Technology*, 18:476-486.

Jenkins, E. W. (1994). Public understanding of science and science education for action. *Journal of Curriculum Studies*, 26, 601-612.

Kind. P., Jones, K., & Barmby, P. (2007). Developing Attitudes towards Science Measures. *International Journal of Science Education*, 29(7), 871-893.

Klopher, L. E. (1971) Evaluation of learning in science. In B.S. Bloom, J.T. Hastings & G. F. Madaus (Eds.), *Handbook of formative and summative evaluation of student learning* (London: McGraw-Hill).

Lind, K. K. (1998, February). Science in early childhood: Developing and acquiring fundamental concepts and skills. Paper presented at the Forum on Early Childhood Science, Mathematics, and Technology Education, Washington, DC.

Lowery, L. F. (1966). Development of an attitude measuring instrument for science education. *School science and mathematics*, 66, 494-502.

Martinez-Beck, I., Tout, K., & Halle, T. (2011). Quality measurement in early childhood settings. Paul H. Brookes Publishing Company.

Mantzicopoulos, P., Patrick, H., & Samarapungavan, A. (2008). Young children's motivational beliefs about learning science. *Early Childhood Research Quarterly*, 23(3), 378-394.

Metz, K. E. (2011). Young Children Can Be Sophisticated Scientists. *Phi Delta Kappan*, 92(8), 68-71.

Munby, H. (1983). Thirty studies involving 'Scientific Attitude Inventory': What Confidence can we have in this Instrument? *Journal of Research in Science Teaching,* 20, 141-162.

Murphy, C., & Beggs, J. (2003). Children's perceptions of school science. *School Science Review,* 84(308), 109-115.

Murphy, C., Ambusaidi, A., & Beggs, J. (2006). Middle East meets West: Comparing children's attitudes to school science. *International Journal of Science Education*, 28(4), 405-422.

Neuman, S. B., Newman, E. H., & Dwyer, J. (2011). Educational Effects of a Vocabulary Intervention on Preschoolers' Word Knowledge and Conceptual Development: A Cluster-Randomized Trial. *Reading Research Quarterly*, 46(3), 249-272.

Nunnally, J. C. Psychometric Theory. McGraw-Hill, New York, NY, 1978.

Ormerod, M. B. & Duckworth, D. (1975). *Pupil's attitude to science* (Slough: NFER).

Ong, E., & Ruthven, K. (2009). The effectivness of smart schooling on students' attitudes towards science. *Eurasia Journal of Mathematics, Science & Technology Education*, 5(1), 35-45.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implication. *International Journal of Science Education, 25*(9), 1049–1079.

Pell, T., & Jarvis, T. (2001). Developing attitude to science scales for use with children of ages from five to eleven years. *International Journal of Science Education, 23*(8), 847–862.

Peleg. R. & Tsabari. B. A. (2011). Atom Surprise: Using Theatre in Primary Science Education. *Journal of Science Education and Technology*, 20:508-524.

Piaget, J. (1929). *The Child's Conception of the World*. Routledge, London.

Reid, N. (2006). Thoughts on attitude measurement. *Research in Science & Technological Education, 24*(1), 3–27.

Rillero, P. (2005). Exploring science with young children. *Early Childhood Today*, 19(6), 8-9.

Roskos, K., & Burstein, K. (2011). Assessment of the Design Efficacy of a Preschool Vocabulary Instruction Technique. *Journal of Research in Childhood Education*, 25(3), 268-287.

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind a comparison of Chinese and US preschoolers. *Psychological Science*, 17(1), 74-81.

Sadi, O., & Cakiroglu, J. (2011). Effects of hands-on activity enriched instruction on students' achievement and attitudes towards science. *Journal of Baltic Science Education*, 10(2), 87-97.

Saracho, O. N., & Spodek, B. (2008). Scientific and technological literacy research. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on science and technology in early childhood education*, 1-16. Charlotte, NC: Information Age.

Schibeci, R. A. (1984). Attitudes to science: An update. *Studies in Science Education, 11,* 26–59.

Silver, A., & Rushton, S. B. (2008). The effect of the Horsham Greenpower Goblin Challenge on children's attitudes towards science, engineering and technology. *Education 3-13: International Journal of Primary, Elementary and Early Years* Education, 36(4), 339-350.

Smith, G. T, & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7,*300-308.

Smithers, A. & Robinson, P. (1988). *The growth of mixed A-levels* (Manchester: Department of

Education, University of Manchester).

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). Using multivariate statistics.

Widaman, K. F. (1993) Common factor analysis versus principal component analysis: Differential bias in representing model parameters? Multivariate Behavioral Research, 28 (3), 263-311.

Worth, K., & Grollman, S. (2003). Worms, shadows, and whirlpools. Portsmouth, NH: Heinneman.