

Measuring University Students' Perceived Self-efficacy In Science Communication in Middle and High Schools

Shaohui Chi^{1*}, Xiufeng Liu², Joseph A. Gardella³

¹ Department of Chemistry, East China Normal University, Shanghai, China

²Department of Learning and Instruction, University at Buffalo, State University of New York, USA

³Department of Chemistry, University at Buffalo, State University of New York, USA.

Copyright©2016 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Service learning typically involves university students in teaching and learning activities for middle and high school students, however, measurement of university students' self-efficacy in science communication is still lacking. In this study, an instrument to measure university students' perceived self-efficacy in communicating science to middle and high school students was developed and validated using a sample of 104 university students (19 graduate students and 85 undergraduate students). The rating scale Rasch model and Winsteps computer program were used to analyze the students' responses to pilot and final revised instrument. The results have revealed that the final revised instrument which contains 20 items with four response categories is well-targeted and measures from this instrument are reasonably valid and reliable. Issues associated with using the instrument are also discussed.

Keywords Measurement Instrument, Perceived Self-efficacy, Science Communication, University Students

There has been well-established evidence on the benefits of placing university students in middle and high school classrooms. For example, teachers involved in the GK-12 program have reported increased STEM content knowledge (e.g., Gamse et al. [1]), a use of more effective pedagogical techniques [2], greater access to STEM resources [3], to name just a few. For middle and high school students, in a recent evaluation of the GK-12 program [4], a majority of teachers indicated that the program had positive effects on their students' STEM knowledge and skills. STEM students working in middle and high school classrooms have reported gains as well. In another recent evaluation of the GK-12 program [5], a majority of current and former graduate students indicated that their GK-12 experience benefited their ability to conduct various activities requiring communication, teaching, and teamwork skills. A majority of their college faculty advisors also concurred that the GK-12 program helps their students develop skills in these areas.

While the benefits of GK-12 and similar service learning programs have been reported as described above, measurement of university students' gains using standardized measurement instruments is still lacking. Our study intends to fill this gap. It focuses on university students' perceived self-efficacy in science communication.

1. Introduction

In the US, there is a long history of involving university students in middle and high school science education. A good example is the NSF funding program called Graduate STEM (Science, Technology, Engineering and Mathematics) Fellows in K-12 Education (GK-12). Through interactions with teachers and students in middle and high schools, graduate STEM fellows improve their science communication and teaching skills while enriching STEM content and instruction for their partners. Over the years, the idea has been expanded to placing university students (graduate and undergraduate) in middle and high school classrooms in order to learn science communication, teaching skills, leadership, teamwork, and civic engagement. This form of university student learning has also been called service learning.

2. Literature Review

Science Communication

Science communication has risen globally in importance in recent years [6]. Science communication is cross-disciplinary, involving communication, psychology, education, philosophy, policy and sociology, as well as the 'traditional' sciences such as natural, physical and computational science [7, 8]. Despite of its importance, science communication has no standard definition. Bryant [9] defines science communication as processes by which the culture and knowledge of science are absorbed into the

culture of the wider community. Gilbert and Stockmayer [10] define science communication as a purposive intervention by a driving actor or a group of driving actors to alter the present state of the relationship between sciences and society toward their desired state. Science communication involves following aspects: Awareness including familiarity with new aspects of science; Enjoyment or other affective response; Interest as evidence by voluntary involvement with science or its communication; Opinions -- the forming, reforming, or confirming of science-related attitudes; and Understanding of science -- its content, processes, and social factors [7].

Shannon and Weaver [11] developed the first model of communication. This one-way science communication model is linear; its aim is for a source to transmit a message to a "receiver" without distortion. Applied to science communication, this model depicts communication as a one-way flow from science to its public and implies a passive public [12] and fails to take into consideration of more complex communication activities, such as feedback from the receiver to the sender [6]. Bryant [9] noted that many scientists hold the idea that knowledge flows like water down a pipe, i.e., from one brain to another without undergoing change. Gilbert and Stockmayer [10] argued that the message about science to be sent always needs to be modified and different receivers may decode the same message in different ways according to their own understandings and thoughts. People learn best when facts and theories have meaning in their personal lives [13].

Currently, more sophisticated two-way models that consider constant feedback in both coding and decoding processes are available [10]. As in the example shown in Figure 1, which was developed by Wood [14], communication is regarded as both interactive and two-way. A contextual model depicts communication as a two-way flow between science and its public and implies an active public; its central focus is not the state of science, but the situation of the public [12].

Specifically for science communication in schools, Bowater and Yeoman[6] propose that a school science communication event should be more structured, fit within a timetabled lesson, and accept that not all kids will be interested in or want to do what have been planned. They suggest that one should ensure that he or she tailors the information to suit the school audience and build upon their existing knowledge. They suggest some steps for people who are planning a school science communication event, such as "think about your audience"; "decide on the subject matter, the aim and objective(s) and how you will deliver it"; and "check the National Curriculum" [6].

Self-efficacy

According to Bandura[15], self-efficacy is a person's belief in one's capabilities to organize and execute the courses of action required to produce certain attainments, and people will not attempt to do things if they do not believe they can produce certain results[15,16]. In other words,

self-efficacy can affect the initiation of behavior, the amount of effort expended and the persistence of behavior in spite of challenges and negative experiences [17]. Other researchers have reached the same conclusion. Self-efficacy not only affects one's cognitive, motivational and affective processes [18] but also determines how the person approaches tasks and responds to set-backs and what the person will do with the skills and knowledge he/she has [19]. The more students succeed, the more they believe they can succeed (self-efficacy), and therefore, the more they do succeed [20].

Self-efficacy is a context-specific rather than a stable characteristic trait. It is therefore thought to have a direct effect on performance in specific contexts. Self-efficacy judgment varies based on the level of skill and perseverance required to achieve a given task in a given context [17,21-23]. Ormrod [24] pointed out that, while self-efficacy is similar to self-concept or self-esteem, an important distinction for self-efficacy is that it is domain, task, or situation specific. Examples provided by Salas [20] is that a teacher may have a strong sense of self-efficacy in teaching mathematics, but weaker self-efficacy in teaching English; or a student may have high self-efficacy when performing mathematics skills, but a low self-efficacy in language arts. Self-efficacy is related to perceived specific abilities rather than generalized self-beliefs[25]. Bursal and Yigit[26] (2012) proposed that self-efficacy beliefs should be extended to specific subject areas since they are context and subject matter dependent.

Over the past decades, many scholars have studied self-efficacy in educational settings, they have found a great influence of self-efficacy on teaching and learning processes (Armor et al., 1976; [22,26,28-38]. Bandura [39] pointed out that educational activities can influence a person's self-efficacy and, therefore, that these activities should utilize methods which can increase self-efficacy. Jones [18] found that self-efficacy can be developed through experience, for example, when one sees that someone like himself/herself succeeds following sustained effort, he/she will believe he/she can succeed too. Other research studies have demonstrated that when training for a specific skill, high self-efficacy is positively correlated with performance [19,40]. Dellinger[21] et al. proposed that self-efficacy be represented in a causal model of interactions among self and society, internal personal factors, and the external environment as reciprocating factors[21]. They argued that internal personal factors (cognitive, affective and biological events) and the external environment influence behaviors, while the environment is impacted by behaviors and personal factors, and personal factors are impacted by behaviors and the environment.

In summary, Communication is essentially as much a matter of listening as it is of talking and, to be effective, each party must have some understanding of the other: "To be effective with any audience, communication must be an interactive process..." [43]. In order to engage the audience, science communicators must identify audience's preconceptions or alternative conceptions of science. The process of participation and engagement in science is a

contextual one [44].

Accordingly, in this study, university students' perceived self-efficacy in science communication was defined as university students' beliefs in their capabilities to help middle and high school students understand science. In our study, science communication is not just about university students' knowledge and understanding of science; it is also about their knowledge of their audience, namely middle and high school students. Specifically, we intend to develop a standardized instrument for measuring university students' perceived self-efficacy in communicating science to middle and high school students. The specific research questions are:

1. What is the validity evidence for supporting the use of the measurement instrument to measure university students' perceived self-efficacy in communicating science to middle and high school students?
2. What is the reliability evidence for supporting the use of the measurement instrument to measure university students' perceived self-efficacy in communicating science to middle and high school students?

3. Method

Participants

The participants were Eighty-seven university students including sixty-eight undergraduate students, one master student and eighteen doctoral students, most of them were in STEM fields (i.e., biological science, chemistry, geological and earth sciences, geography), who took part in a NSF-funded project over three years (2011-2013). Due to IRB protocol, no information on students' ages, gender, racial identities, etc. was collected. These students were assigned to go to local middle and high schools every week to work with students and teachers in science by engaging in such activities as assisting teachers in teaching lessons and find relevant resources, helping students understand science and leading small group activities with students in or after class, etc. Those students completed the pilot instrument after they had completed at least one semester placement in middle and high schools from 2011-2013. Seventeen additional undergraduate students completed the revised instrument after they had completed their placement in local middle and high schools in Dec. 2013.

Procedure

The development of the instrument of university students' perceived self-efficacy in science communication followed a construct modeling approach [45-46]. The construct modeling approach to developing a measurement instrument starts with a clearly defined construct, which "precipitates an idea or a concept that is the theoretical object of our interest in the respondent..." [46], operationalized by progress variables. Assessment tasks are then derived from the defined progress variables, and data collected from

pilot-testing and field-testing are used to examine the fit between the progress variables and data using Rasch modeling[47-48].

In our study, the construct of science communication self-efficacy was defined as the university students' beliefs in their capabilities to help middle and high school students understand science. We used a Likert-scale [49] type question format. Using response scales to collect attitude data has a long history in science education, for each Likert-scale item, respondents are asked to specify their levels of agreement to a given statement, usually expressed in a format such as: strongly disagree, disagree, neutral, agree, strongly agree[47]. The pilot measurement instrument contained 20 items with five response categories to describe respondents' levels of self-efficacy in communicating science. Response categories were coded as 1 through 5 in an ordinal scale: 1-Nothing, 2-Very Little, 3-Some Influence, 4-Quite a Bit, and 5-A Great Deal. The items related to three major aspects of the progress variable on science communication to middle and high school students: understanding students, developing science content, and explain the content.

Data Analysis

Student responses to the 20-item pilot measurement instrument were analyzed using the rating scale Rasch model [50]. In the past 30 years, Rasch measurement has been increasingly used in a wide variety of disciplines[51], and is becoming the convention for developing quality measurement instruments in all social sciences [52]. Based on item response theory (IRT) model, the Rasch model, as a one-parameter logistic mode, provides information of construct validity by fit statistic [52]. When there is good model-data-fit, measures produced by the instrument are interval, the interval scale measures have precise measurement errors for both individual items and subjects, allowing for inferential statistical analyses to be conducted with more power. Compared with classical test theory (CTT), Rasch models have several advantages [53], i.e., while the Classical Test Theory (CTT) analyses attach less importance to the functioning of specific items [54]. Rasch analyses can identify poor patterns of items and person performance, i.e., inform how well the model fits the data, and detect weak, biased, redundant items [55-56]. Embretson and Reise[57] also state that IRT models have four advantages over the CTT model: (a) an IRT trait level estimate can be derived from any items for which properties are known, (b) item properties are directly linked to test behaviors, and (c) the independent variables, trait level and item properties, can be estimated separately without additional data (p. 61).

In terms of reliability, we used the person separation index and the item separation index provided by Winsteps to evaluate the reliability of measures. The person separation index is an estimate of the adjusted person standard deviation divided by the average measurement error, indicates how well the instrument can discriminate persons on the

measured variable. The item separation index indicates an estimate in standard error units of the spread of separation of items along the measurement construct [58]. The reliability separation index greater than two is considered adequate [47].

In regard to the substantive aspect of validity, our evaluation of the instrument focused on item quality proposed by Liu and Boone's [51] framework of validity evidence. According to Liu and Boone [51], "if assessment data fit the Rasch model well, then there is evidence to claim that the originally hypothesized dimension or construct exists, and is assessed by the instrument, thus providing evidence for content and construct validity"[51]. We examined item quality indices (i.e., the mean square residual, the standardized mean square residual) for each item from the rating scale model as implemented in Winsteps computer program [59]. The mean square residual (MNSQ) and the standardized mean square residual (ZSTD) are typically used as the fit indicators to examine how well each item accords with the Rasch unidimensional model. Item MNSQ has an expected value of 1.0 and a range from zero to infinity. Mean-squares greater than 1.0 indicate the data are less predictable than the model expects (underfit), i.e., a mean-square of 1.4 indicates that there is 40% more randomness in the data than modeled. Mean-squares less than 1.0 indicate fits better than expected (overfit), i.e., a mean-square of 0.6 indicates a 40% deficiency in Rasch-model-predicted randomness, which implies $100 \cdot (1-0.6)/0.6 = 67\%$ more ambiguity in the inferred measure than modeled (high discrimination). Based on Linacre's suggestion (Linacre, 2010), items fit the model when their MNSQs fall within the range of 0.6 to 1.4 (for rating scale). ZSTD values are within the range of -2 to +2 (Liu, 2010) when there is a good fit; a positive z-residual indicates that responses are worse than expected; a negative z-residual indicates that responses are better than expected [60]. Item-measure correlation (point-measure correlation/PTMEA) were also examined in this study, zero or negative point-measure correlation indicates a rating scale with reversed direction [61].

The Rasch model constructs a one-dimensional measurement system regardless of the facts that empirical data are always more than on latent dimension [62]. In this study, PCA (principal component analysis) was applied to standardized residuals to identify possible dimensions existing in the scale [63]. A variance greater than or equal to 50% for the Rasch dimension can be considered good [64], and scale unidimensionality also can be assumed if the second dimension (first contrast) has the strength of less than 3 items (in terms of eigenvalues) and the unexplained variance by the first contrast is less than 5% [63]. However, there is no agreement on criteria for representing the existence of a secondary dimension when working with standardized residual-based PCA [61,65-69].

We also use Rasch analyses to verify and improve the functioning of rating scale categorization [70], because how

effectively an instrument's rating scale structure represents a construct is a substantive aspect of validity evidence [71], and effective structure increases the accuracy and precision of the resulting measures, the likelihood of measure stability, and related inferences for future samples [70,72].

4. Results

Pilot-study Item and person separation and reliability

Based on the analysis of the pilot instrument, item separation was 3.33 (reliability=0.92) and person separation was 2.56 (reliability=0.87), both were acceptable. The mean of the infit mean squares (MNSQ) at 1.01 and the outfit mean squares (MNSQ) at 0.99 were very close to the expected value of one. The mean infit ZSTD and outfit ZSTD were both inside the conventionally acceptable range of -2 to +2.

Person Ability and Item Difficulty Measures

From Figure 2, we can see that the Wright map of items and subjects showed that students' self-efficacy had a wide range of variation (person ability measures ranged from -1.34 to 3.33 logits). However, the item difficulty measures ranged from -0.68 to 0.84 logits, narrower than the range of person ability measures. Most items gathered along the middle to lower end of the subjects' communication efficacy range, no item was available for higher science communication efficacy subjects. There are three items at the low levels of the scale, and only one student fell below them, suggesting that those three items need to be improved or removed. Item 16, "Lead small group activities/discussions with students after school or during weekends" (0.84 logits), item 14, "Facilitate out-of-school science learning activities" (0.75 logits), item 10, "Develop out-of-school science learning activities" (0.74 logits), item 19, "Tutor students after school or during weekends" (0.72 logits) were the hardest four items to endorse, indicating that respondents feel relatively less self-efficacy in the aspect of explain the science content during weekends or out of school. The above findings suggested that the items of pilot instrument as a whole were relatively easy for those respondents, thus there was a need for addition of more difficult items for higher efficacy students.

Fit Statistics for Items

The purpose of the fit statistics is to aid in measurement quality control, to identify which data meet Rasch model specifications which don't (RMT <http://www.rasch.org/rmt/rmt103a.htm>). Inspection of the fit statistics for all pilot 20 items (seen in Table 1), 17 of the 20 items had infit and outfit MNSQs within the acceptable range of 0.6 to 1.4, with exceptions of item 2 (infit MNSQ=0.47 and outfit MNSQ=0.49), item 5 (infit MNSQ=1.43 and outfit MNSQ=1.51), and item 19 (infit MNSQ=1.44 and outfit MNSQ=1.45). Six of the 20 items had infit and outfit ZSTD values out of the expected range of -2 to +2: item 1 (infit ZSTD=-2.4 and outfit ZSTD=-2.3), item 2 (infit ZSTD=

-4.2 and outfit ZSTD= -4.1), item 3 (infit ZSTD= -2.6 and outfit ZSTD= -2.0), item 5 (infit ZSTD= 2.5 and outfit ZSTD= 2.9), item 16 (infit ZSTD= 2.2 and outfit ZSTD= 2.2), item 19 (infit ZSTD= 2.8 and outfit ZSTD= 2.8). None of the 20 items had a zero or negative point-measure correlation (PTMEA); all of the point-measure correlations had values ranging from 0.25 to 0.73, which indicated that all of the 20 items contributed to the measurement of students' science communication efficacy.

Unidimensionality

From Table 1, we see factor loadings of the 20 items ranged from -0.56 to 0.68. Items 11, 12, 13, 15, 17, 18, 10 had the contrast loadings over 0.40, and items 3, 4, 6, 7 produced factor loadings of less than -.50, suggesting that they might measure additional dimension. Total variance accounted for 39.1%, the first component had an eigenvalue of 3.5, representing 17.6% of the total variance, below the expected norm. Eigenvalue of components 2 to 5 was 3.2, 2.8, 1.9, and 1.2, respectively, and the proportion of total variance accounted by component 2 to 5 was 9.9%, 8.4%, 5.7%, and 3.7%, indicating that unidimensionality of items was not ideal.

Rating Scale Category Structure

Our item category frequencies had a good spread, meeting the expectations [74]. The measure for category 1 was -2.80, meaning that the average agreeability estimate for persons answering 1 across all items was -2.80 logits. For categories of 2, 3, 4, 5, the category agreeability estimate was -1.30 logits, -0.13 logits, 1.25 logits, and 3.10 logits, respectively, meeting the requirement of the rating scale design, which was increasing monotonically with category.

The step calibration of the 20 items increased monotonically by 0.51 logits, 1.30 logits, and 1.37 logits; however, the category threshold between category 2 ("very little") and category 3 ("some influence) was too close for respondents to differentiate, suggesting that the respondents did not reliably distinguish between the two categories[58].

Item and Instrument Revisions

Based on the Rasch analysis results of the pilot-study instrument, a number of improvements were made to the instruments. Specifically, in order to accurately measure the university students' perceived self-efficacy in science communication of persons with the highest ability level, we added four new items: new item 17, "Explain a difficult science concept to students", new item 18, "Explain current research to teachers", new item 19, "Facilitate student learning in museums", new item 20, "Explain science to parents".

Four pilot items had similar measures and had more or less poor fit indicators: item "understand middle and high school students' science background knowledge" (-0.41 logits, infit ZSTD= -2.4 and outfit ZSTD= -2.3), item "understand middle and high school students' interest in science" (-0.43 logits, infit ZSTD= -4.2 and outfit ZSTD= -4.1, infit

MNSQ= 0.47 and outfit MNSQ= 0.49), item "Understand middle and high school students' social and cultural backgrounds" (-0.34 logits), item "Understand middle and high school students' attention span" (-0.53 logits, infit ZSTD= 2.5 and outfit ZSTD= 2.9, infit MNSQ= 1.43 and outfit MNSQ= 1.51). In terms of the analysis and considered that "Understand middle and high school students' social and cultural backgrounds" and "Understand middle and high school students' attention span" (one of the easiest items to endorse in pilot study) may less related to the measured construct, therefore, we removed these two items from the instrument. We also removed item "Lead small group activities/discussions with students after school or during weekends"(infit ZSTD= 2.2 and outfit ZSTD= 2.2) and item "Tutor students after school or during weekends" (infit ZSTD= 2.8 and outfit ZSTD= 2.8, infit MNSQ=1.44, outfit MNSQ=1.45), which were the hardest items to endorse in pilot study, because they not only poor fit the model but also pertained to "weekends" activities that were not central to the measured construct.

According to Linacre[70], "For a five category rating scale, advances of at least 1.0 logits between step calibrations are needed in order for that scale to be equivalent to four dichotomies...when the advance is less than 1.0 logits ...redefining the categories to have wider substantive meaning or combining categories may be indicated" [70]. Other researchers also report that collapsing one or two categories will increase the test reliability[72, 74], i.e., Stone and Wright [75] found in their survey of perceived fear, combining five categories into three increases the test reliability. Therefore, we collapsed the rating scale categories from five to four. The new categories became: 1—Little, 2—Some, 3—Quite a bit, and 4—A Great Deal.

Field-testing

The revised instrument included again 20 items which were then responded by 17 university students. Responses by the 17 university students were combined with the responses by the former 87 university students from the pilot study by the following recoding: 1 was coded as 1, 2 and 3 were coded as 2, 4 as 3, and 5 as 4. The combined responses were then submitted to Rasch analysis again. The findings reported next are based on this analysis.

Resulted from the revised instrument, the person separation index was 2.77, with an equivalent Cronbach's reliability coefficient (α value) of 0.88. Item separation index was 2.94, and its corresponding Cronbach's α value was 0.90, indicating reliable item and person estimation. Rasch measurement also produces an SEM as an additional measure of reliability for each individual person and item measure. Persons and items with measures closer to their means have smaller SEMs than those further from the means, SEM values for persons and items were small, ranging from 0.14 to 0.33.

Figure 3 presents the Wright map of the revised instrument, we can see that university students' perceived self-efficacy measures have a wider range of variation from -2.33 logits to 5.92 logits, while the revised item measures

also have a wider range from -0.97 logits to 1.23 logits. The first two most difficult items (item 20, item 19) were the new items (1.23 logits, 1.12 logits), and item 17 (0.39 logits) and item 18 (0.60 logits) were both above the mean of the items, indicating that the four new items were relatively difficult items just as intended. However, there was still one gap located near two standard deviations from the mean of the items; fifteen university students had a lower perceived self-efficacy than any item could assess. Another gap existed at the top of the continuum, where 14 higher perceived self-efficacy university students were in that gap.

Table 2 presents fit statistics for the final 20 items in the revised instrument. We can see that infit MNSQs ranged from 0.65 to 1.29 whereas the outfit MNSQs ranged from 0.69 to 1.31; both were regarded as being acceptable. Infit ZSTDs and outfit ZSTDs all ranged from -2.0 to +2.0 with the exception of item 2 (infit ZSTD = -3.0 and outfit ZSTD = -2.5), item 6 (infit ZSTD = 1.8 and outfit ZSTD = 2.2). All the items exhibited strong positive point-measure correlations (PTMEA) ranging from 0.50 to 0.70.

Measures resulted from the revised measurement accounted for 43.9% of total variance, though 4% higher than pilot measurement, yet still below the expected norm. The first component had an eigenvalue of 3.2, representing 16.0% of the total variance. Eigenvalue of components 2 to 5 was 2.6, 2.1, 1.7, and 1.5, respectively, and the proportion of total variance accounted by component 2 to 5 was 9.9%, 8.4%, 5.7%, and 3.7%, indicating that unidimensionality of items was not ideal. From Table 2, we see factor loadings of the 20 items ranged from -0.62 to 0.67. Items 9, 10, 11, 14 had the contrast loadings over 0.40, and items 3, 5 produced factor loadings of less than -.50, suggesting that they still might measure additional dimension.

Table 3 presents the category structure statistics. As shown in Table 3, with four categories instead of five, each category count satisfied the criterion for minimum counts of 10 observations [70]. The average category measures were ordered and increased monotonically from -1.01 logits to 1.60 logits. The outfit MNSQ ranged from 0.96 logits to 1.02 logits, indicating expected category usage [70]. In addition, the category threshold calibrations increased monotonically with categories and the distances were all more than 1.1 logits, meeting the guidelines given by Linacre [70]. Inspecting the category probability curves (see Figure 4), we see that each category represented a distinct region of the underlying construct, thus, collapsing category 1 and 2 had indeed improved our rating scale diagnostics.

5. Discussion

The purpose of this study was to develop a standardized instrument for measuring university students' perceived self-efficacy in communicating science. In order to evaluate the validity and reliability of the instrument, we first conducted a pilot study and examined the person separation index and item separation index, person ability and item difficulty measures, item quality indices, unidimensionality,

and the functioning of rating scale categorization of the pilot instrument using Rasch model analysis. Then, based on the analyses of the results, we improved the pilot instrument and examined the revised instrument.

From the above presented findings, our revised instrument appeared to be highly reliable as indicated by the Rasch reliability statistics. Overall, items fit the Rasch model well, suggesting that there is evidence for the construct validity of the revised instrument measures. Examination of the person-item map distribution of revealed that the revised instrument item difficulty measures were better than before, but still narrower than the range of person ability measures, with absence of items at the high end of the scale, suggesting more difficult items need to be added.

Item 20, "Explain science to parents" (1.23 logits) and item 19, "Facilitate student learning in museums" (1.12 logits) were ranked top on the difficulty column. Parents have a strong influence on children development; they not only influence children's in-school achievement but also make decisions about children's out-of-school activities [76]. Communicating science with parents is a good way to help their children better understand science but may also be a hard way that needs years of teaching experiences and skills. As for university students, mostly have little teaching experiences; it can be expected that they feel relatively less self-efficacy on "Explain science to parents". Middle and high school students enjoy visits to museums and can also benefit a lot from visiting, yet in order to maximize the benefits of the visit, more work need to be done by our teachers, for example, Carr [77] provided some useful guidelines for facilitate student learning in museums: (a) children in museums should have opportunities to interpret open questions about the meaning of evidence. (b) children in museums should have opportunities to construct knowledge, rather than receive it. (c) children in museums should have sustained encounters with process, ambiguity, collaboration, and mystery, encounters leading to grounded knowledge of how thinking happens...[77]. Therefore, it needs to pay more attention and think carefully about what museum, especially science museum has to offer and how it can be related to what students are learning about science in order to "Facilitate student learning in museums", it also would be the reason to explain why most of our respondents as university students had less efficacy on this item. According to Bandura [40], Self-efficacy beliefs come from four main sources: (1) mastery experiences, (2) vicarious experiences, (3) verbal persuasion, and (4) physiological indexes. Among these sources, mastery experiences were claimed to be the most important self-efficacy source, the results above are consistent with the claim.

Although unidimensionality of the revised instrument is still less ideal, yet it is common in the literature involving Rasch analysis that reported variance accounted for by Rasch measures based on PCA is less than 50% [63,79,80] (Cervellione et al., 2009; , and our variance accounted for 43.9% by Rasch measures in this study is decent and not unusual.

After collapsing two categories, the four categories provide better functioning of the scale. Actually the issue of the preferred number of responses on a Likert scale has been discussed much on research methodology [82], some researchers argued that in a five-categorized Likert scale there is a middle box representing a neutral category and the respondents tend to choose the middle box for various reasons [83]. Since in an attitude Likert scale, we do not know exactly where the attitudes turn from slightly positive to slightly negative [81], it may be better use even number categories for attitude scale.

Overall, despite the aforementioned limitations, the results suggest that the revised instrument of university students' perceived self-efficacy in science communication with the

new 20-items is well-targeted at the university students. Measures from this instrument are reasonably valid and reliable, thus are appropriate for assessing university students' perceived self-efficacy in science communication.

Acknowledgements

The materials reported in this paper are based upon work supported by the National Science Foundation under Grant No. DUE-1102998 and project title of the University at Buffalo/Buffalo Public Schools Interdisciplinary Science and Engineering Partnership.

Appendix

Table 1. The Original Item Statements and Statistics

Item	Statement	Measure	Infit		Outfit		PTMEA	Loading
			MNSQ	ZSTD	MNSQ	ZSTD		
1	Understand middle and high school students' science background knowledge	-0.41	0.66	-2.4	0.68	-2.3	0.58	-0.25
2	Understand middle and high school students' interest in science	-0.43	0.47	-4.2	0.49	-4.1	0.64	-0.22
3	Understand middle and high school students' cognitive abilities	-0.10	0.65	-2.6	0.72	-2.0	0.51	-0.55
4	Understand middle and high school students' social and cultural backgrounds	-0.34	1.26	1.6	1.29	1.8	0.32	-0.56
5	Understand middle and high school students' attention span	-0.53	1.43	2.5	1.51	2.9	0.25	-0.27
6	Decide what science topics are appropriate to students	0.04	0.83	-1.1	0.81	-1.3	0.55	-0.53
7	Decide how much science content is appropriate to students	0.36	0.90	-0.6	0.95	-0.3	0.45	-0.52
8	Help teachers find relevant resources (e.g., science activities)	0.06	1.20	1.3	1.16	1.1	0.52	-0.31
9	Develop science labs	0.58	1.27	1.8	1.27	1.8	0.57	0.07
10	Develop out-of-school science learning activities	0.74	0.95	-0.3	0.95	-0.3	0.59	-0.19
11	Assist teachers in teaching lessons	-0.43	1.04	0.3	1.04	0.3	0.62	0.49
12	Assist teachers in conducting labs	-0.68	1.02	0.2	0.96	-0.2	0.64	0.59
13	Teach science labs to students	-0.15	0.99	0.0	0.96	-0.2	0.65	0.68
14	Facilitate out-of-school science learning activities	0.75	0.73	-2.0	0.74	-1.9	0.71	-0.10
15	Lead small group activities/discussions with students in class	-0.57	1.17	1.1	1.07	0.5	0.56	0.46
16	Lead small group activities/discussions with students after school or during weekends	0.84	1.33	2.2	1.33	2.2	0.53	-0.09
17	Demonstrate scientific content, procedures, tools, or techniques to students	-0.41	0.92	-0.5	0.86	-0.9	0.64	0.67
18	Teach lessons or give lectures to students in class	0.16	0.87	-0.9	0.86	-0.9	0.73	0.43
19	Tutor students after school or during weekends	0.72	1.44	2.8	1.45	2.8	0.51	-0.01
20	Explain a difficult science concept to students	0.20	0.76	-1.6	0.76	-1.7	0.65	0.36

Table 2. The Revised Item Statements and Statistics

Item	Statement	Measure	Infit		Outfit		PTMEA	Loading
			MNSQ	ZSTD	MNSQ	ZSTD		
1	Understand middle and high school students' science background knowledge	-0.54	0.81	-1.5	0.82	-1.3	0.61	-0.20
2	Understand middle and high school students' interest in science	-0.54	0.65	-3.0	0.69	-2.5	0.62	-0.19
3	Understand middle and high school students' cognitive abilities	-0.20	0.94	-0.5	1.02	0.2	0.50	-0.51
4	Decide what science topics are appropriate to students	-0.11	0.94	-0.4	0.97	-0.2	0.60	-0.62
5	Decide how much science content is appropriate to students	0.33	1.12	1.0	1.19	1.4	0.50	-0.54
6	Help teachers find relevant resources (e.g., science activities)	-0.15	1.25	1.8	1.31	2.2	0.57	-0.27
7	Develop science labs	0.42	1.24	1.8	1.21	1.6	0.65	0.01
8	Develop out-of-school science learning activities	0.73	1.12	1.0	1.08	0.6	0.62	-0.33
9	Assist teachers in teaching lessons	-0.70	1.17	1.3	1.18	1.3	0.56	0.56
10	Assist teachers in conducting labs	-0.97	1.08	0.6	1.10	0.7	0.61	0.62
11	Teach science labs to students	-0.34	0.97	-0.2	-0.93	-0.5	0.68	0.67
12	Facilitate out-of-school science learning activities	0.66	0.88	-0.9	0.87	-1.0	0.69	-0.36
13	Lead small group activities/discussions with students in class	-0.71	1.14	1.1	1.09	0.7	0.55	0.32
14	Demonstrate scientific content, procedures, tools, or techniques to students	-0.68	0.92	-0.6	0.87	-0.9	0.65	0.61
15	Teach lessons or give lectures to students in class	-0.09	0.90	-0.8	0.90	-0.7	0.70	0.33
16	Explain a difficult science concept to students	-0.45	0.77	-1.9	0.76	-1.9	0.69	0.14
17	Relate current research to K-12 curriculum	0.39	1.07	0.3	1.03	0.2	0.64	-0.33
18	Explain current research to teachers	0.60	1.04	0.2	0.97	0.0	0.65	0.07
19	Facilitate student learning in museums	1.12	1.29	1.0	1.20	0.7	0.66	-0.02
20	Explain science to parents	1.23	1.23	0.8	1.29	1.0	0.60	-0.22

Table 3. Summary of Rating Scale

Rating Scale Category	Observed Count	Observed%	Average Measure	Outfit MNSQ	Step Calibrations
1=None	203	12	-1.01	1.02	NONE
2=Some	482	28	-0.17	0.96	-1.46
3=Quite a bit	631	36	0.58	1.05	-0.07
4=A great deal	420	24	1.60	1.00	1.53

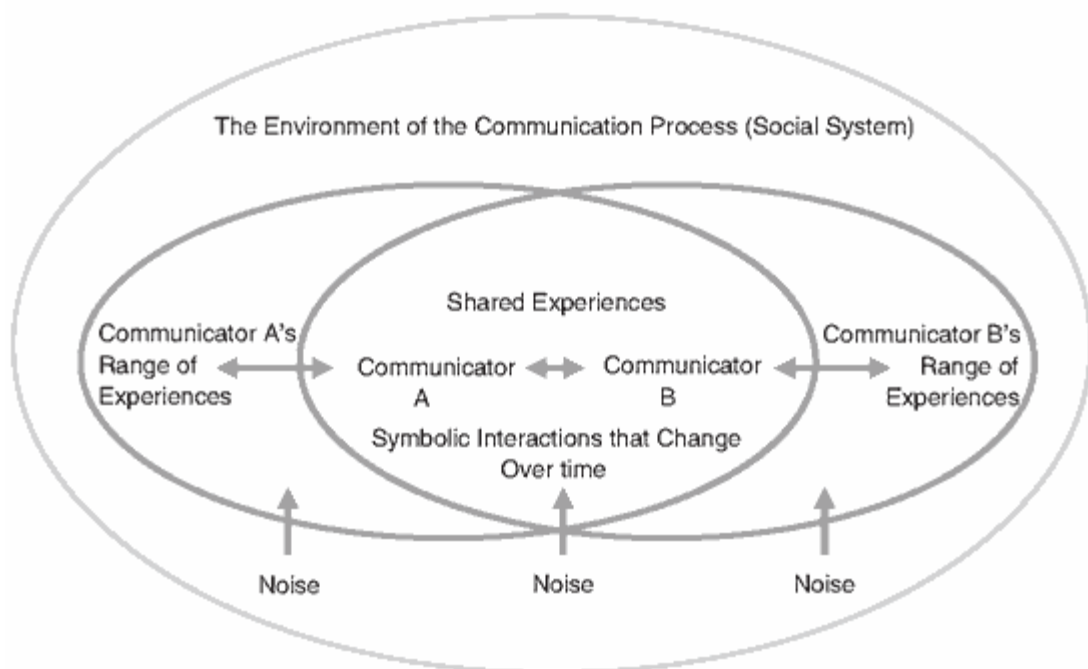


Figure 1. Transaction model of communication from Wood, 2003 (adapted by Bowater & Yeoman, 2012)

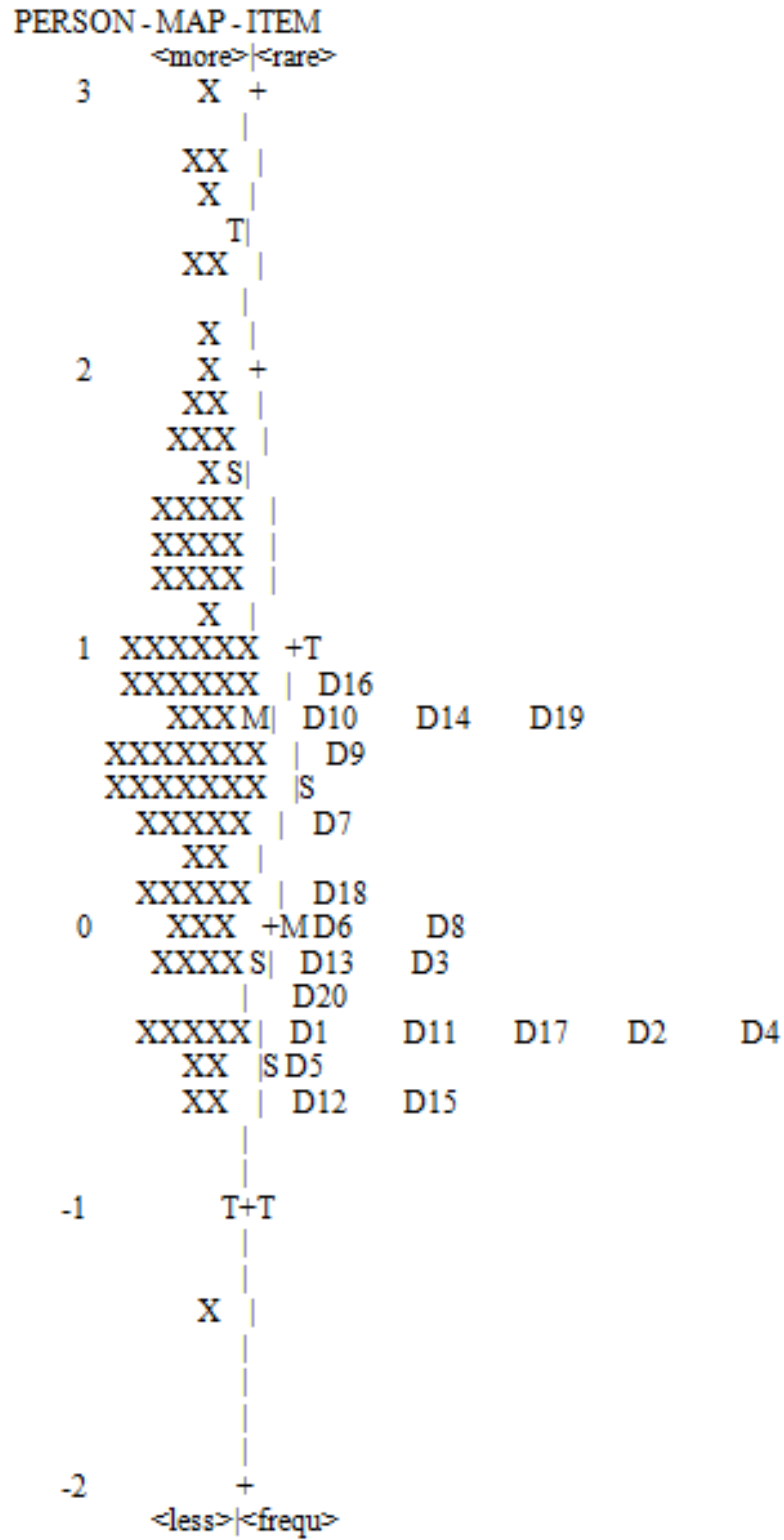


Figure 2. Wright Map of Person-Item distribution (pilot study)

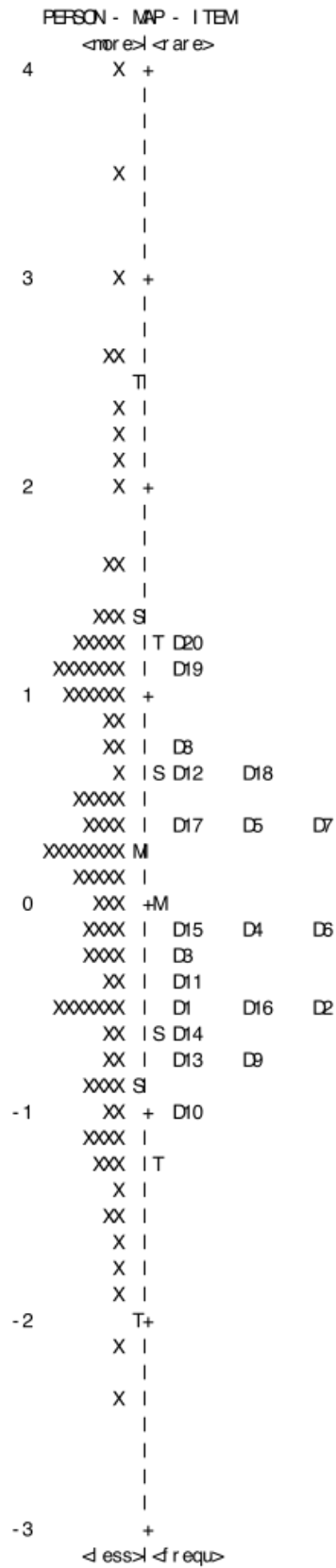


Figure 3. Wright Map of Person-Item distribution (field study)

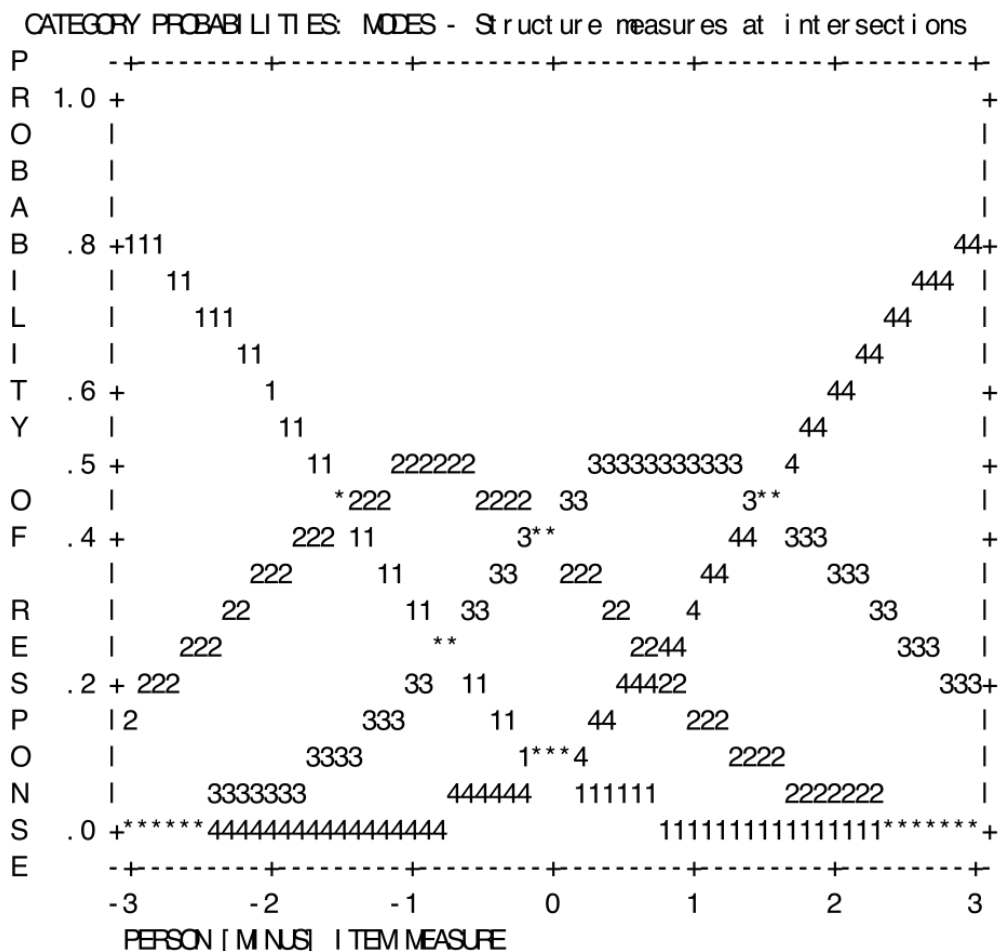


Figure 4. Category structure probabilities curves

REFERENCES

[1] Gamse, B., et al. "Evaluation of the National Science Foundation's GK-12 program." Cambridge, MA: Abt Associates, 2010

[2] Huziak - Clark T, Van Hook S J, Nummerger - Haag J, et al. Using Inquiry to Improve Pedagogy through K - 12/University Partnerships[J]. School Science and Mathematics, 107(8): 311-324, 2007

[3] Moskal, B., Skokan, C., Kosbar, L., Dean, A., Westland, C., Barker, H., Tafoya, J. (2007). K-12 outreach: Identifying the broader impacts of four outreach projects. Journal of Engineering Education, 96(3), 173-189.

[4] Abt Associates Evaluation of the National Science Foundation's GK-12 Program. Arlington, VA,2010.

[5] NSF.(2010) https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5337

[6] Bowater, L., & Yeoman, K. (2012). Science Communication: A Practical Guide for Scientists. John Wiley & Sons.

[7] Burns, T. W., & O'connor, D. J. M., Stockmayer. S. (2003). Science communication: a contemporary definition. Public Understanding of Science 12, 183-202

[8] Mulder, H. A., Longnecker, N., & Davis, L. S. The state of science communication programs at universities around the world. Science Communication, 30(2), 277-287,2008.

[9] Bryant, C. Does Australia need a more effective policy of Science Communication?. International Journal for Parasitology, 33(4), 357-361,2003.

[10] Gilbert, J. K., Lewenstein, B. V., & Stockmayer, S. M. (Eds.). Communication and Engagement with Science and Technology: Issues and Dilemmas A Reader in Science Communication. Routledge,2012.

[11] Shannon, C. E., & Weaver, W. The mathematical theory of communication (Urbana, IL. University of Illinois Press, 19(7), 1,1949.

[12] Gross, A. G. The roles of rhetoric in the public understanding of science. Public Understanding of Science, 3(1), 3-23,1994.

[13] Kahlor, L., & Stout, P. (Eds.). Understanding and Communicating Science. Routledge,2009.

[14] Wood, J. T. Communication in our lives,3rd edn.Thomson,2003.

[15] Bandura, A. Perceived self-efficacy in the exercise of control over AIDS infection. Evaluation and program planning, 13(1), 9-17,1990.

- [16] Bandura, A., ed. *Self-efficacy: The exercise of control*. New York: W.H. Freeman, 1997.
- [17] Bandura, A. Self-efficacy. *Toward a unifying theory of behavioral change*. *Psychological Review*, 84, 191–215, 1997.
- [18] Jones, A. . Developing a measurement tool for assessing physiotherapy students' self-efficacy: a pilot study. *Assessment & Evaluation In Higher Education*, 37(3), 369-377, 2012.
- [19] Pajares, F.. *Assessing Self-Efficacy Beliefs and Academic Outcomes: The Case for Specificity and Correspondence*. 1996.
- [20] Salas, C. S. . *Measuring self-efficacy of teachers using Multiple Intelligences teacher training (Doctoral dissertation, CAPELLA UNIVERSITY)*. 2009.
- [21] Dellinger, A. B., Bobbett, J. J., Olivier, D. F., & Ellett, C. D. . *Measuring teachers' self-efficacy beliefs: Development and use of the TEBS-Self*. *Teaching and Teacher Education*, 24(3), 751-766, 2008.
- [22] Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K.. *Teacher efficacy: Its meaning and measure*. *Review of educational research*, 68(2), 202-248, 1998.
- [23] Woolfolk Hoy, A., & Burke Spero, R.. *Changes in teacher efficacy during the early years of teaching: A comparison of four measures*. *Teaching and teacher education*, 21(4), 343-356, 2005.
- [24] Ormrod, J.. *Learning theory and the educational process (4th ed.)*. Boston: Pearson, 2004.
- [25] Gaffney, A. L. H. . *Measuring Students' Self - Efficacy for Communication*. *International Journal of Art & Design Education*, 30(2), 211-225, 2011.
- [26] Bursal, M., & Yigit, N.. *Pre-Service Science and Technology Teachers' Efficacy Beliefs about Information and Communication Technologies (ICT) Usage and Material Design*. *Educational Sciences: Theory and Practice*, 12(2), 1084-1088, 2012.
- [27] Armor, D., Conroy-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., et al. *Analysis of the school preferred reading programs in selected Los Angeles minority schools (Rep. No. R-2007-LAUSD)*. Santa Monica, CA: RAND, 1976.
- [28] Ashton, P. T. . *Teacher efficacy: A motivational paradigm for effective teacher education*. *Journal of Teacher Education*, 35, 28–32, 1984.
- [29] [29] Ashton, P. T., Webb, R. B., & Doda, N. . *A study of teachers' sense of efficacy (Final Report, National Institute of Education Contract No. 400-79-0075)*. Gainesville: University of Florida. (ERIC Document Reproduction Service No. ED 231 834). 1983.
- [30] Gibson, Sh., & Dembo, M.H. . *Teacher efficacy: A construct validation*. *Journal of Educational Psychology*, 76(4), 569-582, 1984.
- [31] Guskey, T.R., & Passaro, P.D. . *Teacher efficacy: A study of construct dimensions*. *American Educational Research Journal*, 31, 627–643, 1994.
- [32] Humphries C A, Hebert E, Daigle K, et al. *Measurement in Physical Education and Exercise Science. Development of a physical education teaching efficacy scale*, 16(4), 284-299, 2012.
- [33] Roll-Peterson, L. . *Teachers perceived efficacy and the inclusion of a pupil with dyslexia or mild mental retardation: Findings from Sweden*. *Education and Training in Developmental Disabilities*, 43, 174–185, 2008.
- [34] Schunk, D.H. . *Self-efficacy and achievement behaviors*, *Educational Psychology Review*, Vol. 1, No. 3, pp. 173–208, 1989.
- [35] Shim, S., & Ryan, A. . *Changes in self-efficacy, challenge avoidance, and intrinsic value in response to grades: The role of achievement goals*. *Journal of Experimental Education*, 73(4), 333–349. Retrieved April 29, 2008
- [36] Soodak, L. C., & Podell, D. M. . *Teacher efficacy and student problem as factors in special education referral*. *The Journal of Special Education*, 27(1), 66-81, 1993.
- [37] Soodak, L. C., & Podell, D. M. *Teacher efficacy: Toward the understanding of a multi-faceted construct*. *Teaching and Teacher Education*, 12, 401–411, 1996.
- [38] Woolfolk Hoy, A., & Davis, H. A. *Teacher self-efficacy and its influence on the achievement of adolescents*. *Self-efficacy beliefs of adolescents*, 117-137, 2006.
- [39] Bandura, A. “Self-efficacy”, in Ramachandran, V.S. (Ed.), *Encyclopedia of Human Behavior*, Vol. 4, Academic Press, New York, NY, pp. 71-81, 1994.
- [40] Bandura, A. *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall, 1986.
- [41] Block, M. E., Hutzler, Y., Barak, S., & Klavina, A. . *Creation and Validation of the Self-Efficacy Instrument for Physical Education Teacher Education Majors Toward Inclusion*. *Adapted Physical Activity Quarterly*, 30(2), 184-205, 2013.
- [42] Gist, M., Schwoerer, C., & Rosen, B. . *Effects of alternative training methods on self-efficacy and performance in computer software training*. *Journal of Applied Psychology*, 74(6), 884 – 891, 1989.
- [43] Stocklmayer, S. M., Gore, M. M., & Bryant, C. R. (Eds.). . *Science communication in theory and practice (Vol. 14)*. Springer, 2001.
- [44] Falk, J., & Storksdieck, M. . *Using the contextual model of learning to understand visitor learning from a science center exhibition*. *Science Education*, 89(5), 744-778, 2005.
- [45] Wilson, M.. *On choosing a model for measuring*. 2003.
- [46] Wilson, M. . *Constructing measures: An item response modeling approach*. 2005.
- [47] Bond, T. G., & Fox, C. M. . *Applying the Rasch model: Fundamental measurement in the human sciences (second edition)*. Mahwah, NJ: Lawrence Erlbaum, 2007.
- [48] Liu, X. . *Using and developing measurement instruments in science education: A Rasch modeling approach*. Iap, 2010.
- [49] Likert, R. A. . *Technique for the measurement of attitudes*, *Archives of Psychology*, 140, 1932.
- [50] Andrich, D. *Rating formulation for ordered response categories*. *Psychometrika*, 43(4), 561-573, 1978.

- [51] Liu, X., & Boone, W. J. Applications of Rasch measurement in science education. JAM Press.2006.
- [52] Royal, K. D., Ellis, A., Ensslen, A., & Homan, A.. Rating scale optimization in survey research: an application of the Rasch rating scale model. *Journal of Applied Quantitative Methods*, 5(4), 607.2010.
- [53] Sussman, J., Beaujean, A., Worrell, F. C., & Watson, S. . An Analysis of Cross Racial Identity Scale Scores Using Classical Test Theory and Rasch Item Response Models. *Measurement And Evaluation In Counseling And Development*, 46(2), 136-153,2013.
- [54] McDonald, R. P. . Test theory: A unified treatment. Mahwah, NJ: Erlbaum.1999.
- [55] Hula, W.D., Doyle, P.J., McNeil, M.R., Mikolic, J.M. . Rasch modeling of revised Token Test performance: Validity and sensitivity to change. *Journal of Speech-Language-Hearing Research*, 49, 27-46,2006.
- [56] Thissen, D., & Steinberg, L. . Data analysis using item response theory. *Psychological Bulletin*, 104, 385- 395. doi:10.1037/0033-2909.104.3.385.1988.
- [57] Embretson, S.E., & Reise, S.P. Item response theory for psychologists. Mahawah, NJ: Lawrence Erlbaum.2000.
- [58] Kim, DH., Wang, C., Ng, KM. A Rasch Rating Scale Modeling of the Schutte Self-Report Emotional Intelligence Scale in a Sample of International Students. *Assessment*, 17(4), 484-96,2010.
- [59] Linacre, J. M. Winsteps (version 3.72.0) [Computer Software]. Chicago: Winsteps.com.2011.
- [60] Bradley, K. D., Cunningham, J., Haines, T., Mueller, C. E., Royal, K. D., Sampson, S. O., ... & Weber, J. A Constructing and evaluating measures: Applications of the Rasch measurement model.2010.
- [61] Nam, S. K., Yang, E., Lee, S. M., Lee, S. H., & Seol, H. . A psychometric evaluation of the career decision self-efficacy scale with Korean students: a Rasch model approach. *Journal of Career development*, 38(2), 147-166,2011. Dellinger, A. B. Validity and the review of literature. *Research in the Schools*, 12(2), 41-54,2005.
- [62] Linacre, J. M. Detecting Multidimensionality: Which Residual Data-Type Works Best? *Journal of Outcome Measurement*, 2, 266-283,1998.
- [63] Oon, P. T., & Subramaniam, R. Rasch Modelling of a Scale that Explores the Take-Up of Physics Among School Students from the Perspective of Teachers. In *Applications of Rasch measurement in learning environments research* (pp. 119-139). SensePublishers.2011.
- [64] Linacre, J. M. WINSTEPS Rasch measurement computer program, version 3.69.1.8.Chicago, IL: Winsteps.com.2010.
- [65] Andrich, D., Sheridan, B., & Luo, G. .Interpreting RUMM2020 [Monograph].Retrieved February 13, 2008, Online available from <http://www.rummlab.com>
- [66] Raiche, G. Critical eigenvalue sizes in standardized residual principal component analysis. *Rasch Measurement Transaction*, 19, 1012.2005.
- [67] Reckase, M. Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207-230,1979.
- [68] Smith, R. M., & Miao, C. Y. . Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective measurement: theory into practice* Vol. 2., 316-327,1994.).
- [69] Smith, R. M. . A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling*, 3, 25-40,1996.
- [70] Linacre, J. M.. Optimizing rating scale category effectiveness. *J Appl Meas*, 3(1), 85-106,2002.
- [71] Messick, S.. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749,1995.
- [72] Myers, N. D., Feltz, D. L., & Wolfe, E. W. . A confirmatory study of rating scale category effectiveness for the coaching efficacy scale. *Research Quarterly for Exercise and Sport*, 79, 300-311,2008.
- [73] Wolfe, E. W., & Smith Jr, E. V. Instrument development tools and activities or measure validation using Rasch models: part II--validation activities. *Journal of Applied Measurement*, 8(2), 204-234,2006.
- [74] Zhu, W., Updyke, W. F. & Lewandowski C. . Post-Hoc Rasch analysis of optimal categorization of an ordered response scale. *Journal of Outcome Measurement*, 1:4, 286-304,1997.
- [75] Stone M. H. & Wright B.D. Maximizing rating scale information. *Rasch Measurement Transactions*, 8, 3, 386,1994.
- [76] Eccles, J.S. Families, schools, and developing achievement-related motivations and engagement. In J.E. Grusec & P.D. Hastings (Eds.), *handbook of socialization*, New York: Guilford.2007.
- [77] Carr, D. . The promise of cultural institutions. Walnut Creek, CA: Alta Mira Press.2003.
- [78] Fischer, A. R., De Jong, A. E., De Jonge, R., Frewer, L. J., & Nauta, M. J. . Improving food safety in the domestic environment: The need for a transdisciplinary approach. *Risk Analysis*, 25(3), 503-517,2005.
- [79] Higgins, G. E. . Examining the Original Grasmick Scale A Rasch Model Approach. *Criminal Justice and Behavior*, 34(2), 157-178,2007.
- [80] Cervellione K L, Lee Y S, Bonanno G A. Rasch modeling of the self-deception scale of the balanced inventory of desirable responding. *Educational and Psychological Measurement*, 2008.
- [81] Schreiner, C., & Sjøberg, S. . Sowing the seeds of ROSE. Background, Rationale, Questionnaire Development and Data Collection for ROSE (The Relevance of Science Education) - a comparative study of students' views of science and science education .2004.
- [82] Gable, R.K. & Wof, M.B.. Instrument Development in the Affective Domain. *Measuring Attitudes and Values in Corporate and School Settings*. Boston: Kluwer Academic Publishers.1993.
- [83] Ashton, P. T., Olejnik, S., Crocker, L., & McAuliffe, M. (1982). Measurement problems in the study of teachers' sense of efficacy. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

- [84] Ashton, P.T., & Webb, R.B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York: Longman.
- [85] Berman, P. (1977). *Federal Programs Supporting Educational Change, Vol. VII: Factors Affecting Implementation and Continuation*
- [86] Berridge, E. J., Freeth, D., Sharpe, J., & Roberts, C. M. (2007). Bridging the gap: supporting the transition from medical student to practising doctor—a two-week preparation programme after graduation. *Medical teacher*, 29(2-3), 119-127.
- [87] Brouwers, A., & Tomic, W. (2003). A test of the factorial validity of the Teacher Efficacy Scale. *Research in Education*, 69(1), 67-79.
- [88] Clark, M. C., Owen, S. V., & Tholcken, M. A. (2004). Measuring student perceptions of clinical competence. *The Journal of nursing education*, 43(12), 548.
- [89] Coladarci, T., & Fink, D. R. (1995). Correlations among measures of teacher efficacy: Are they measuring the same thing? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.
- [90] Day, C. S., Yeh, A. C., Franko, O., Ramirez, M., & Krupat, E. (2007). Musculoskeletal medicine: an assessment of the attitudes and knowledge of medical students at Harvard Medical School. *Academic Medicine*, 82(5), 452-457.
- [91] DeChenne, S. E., & Enochs, L. (2010). Measuring the Teaching Self-Efficacy of Science, Technology, Engineering, and Math Graduate Teaching Assistants. Online Submission.
- [92] Deemer, S. A., & Minke, K. M. (1999). An investigation of the factor structure of the teacher efficacy scale. *The Journal of Educational Research*, 93(1), 3-10. Denzine, G. M., Cooney, J. B., & McKenzie, R. (2005). Confirmatory factor analysis of the Teacher Efficacy Scale for prospective teachers. *British Journal of Educational Psychology*, 75(4), 689-708.
- [93] Gamse, B., Carter Smith, W., Parsad, A., Dreier, D., Neishi, K., Carney, J., Spader, J. (2010). Evaluation of the National Science Foundation's GK-12 Program. Volume I and II. Cambridge, MA: Abt Associates, Inc; Prepared for the National Science Foundation, Arlington, VA.
- [94] Guskey, T. R. (1981). The relationship of affect toward teaching and teaching self-concept to responsibility for student achievement. *Journal of Social Studies Research*, 5(69-74)
- [95] Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, 37, 137-150
- [96] Huziak-Clark, T., Van Hook, S., Nurnberger-Haag, J., & Ballone-Duran, L. (2007). Using inquiry to improve Pedagogy through K-12/University partnerships. *School Science and Mathematics*, 107(8), 311-324.
- [97] Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998-2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, 23, 21-43
- [98] Moskal, B., Skokan, C., Kosbar, L., Dean, A., Westland, C., Barker, H., Tafoya, J. (2007). K-12 outreach: Identifying the broader impacts of four outreach projects. *Journal of Engineering Education*, 96(3), 173-189.
- [99] Riboh, J., M. Curet, and T. Krummel. (2007). Innovative introduction to surgery in the preclinical years. *American Journal of Surgery* 194: 227-30.
- [100] Roberts, J. K., & Henson, R. K. (2000). Self-Efficacy Teaching and Knowledge Instrument for Science Teachers (SETAKIST): A Proposal for a New Efficacy Instrument.
- [101] Rose, J. S., & Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *Journal of Educational Research*, 74, 185-190.
- [102] Shea, J. A., & Fortna, G. S. (2002). Psychometric methods. In *International handbook of research in medical education* (pp. 97-126). Springer Netherlands.
- [103] Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B. & Rogers, R. W. (1982). The self-efficacy scale: construction and validation, *Psychological Reports*, Vol. 51, No. 2, pp. 663-71.
- [104] Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and teacher education*, 17(7), 783-805. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- [105] Ziman, J. (1991). Public understanding of science. *Science, Technology & Human Values*, 16(1), 99-105.
- [106] Ziman, J. (1992). Not knowing, needing to know, and wanting to know. When science meets the public, 13-20.