Teaching and Learning in Communication Sciences & Disorders

Volume 7 | Issue 1 Article 3

2023

Examining the Item-level Psychometric Properties of the Self-Efficacy Questionnaire (SEQ) for Students in Communication Sciences and Disorders

Sabiha Parveen
Oklahoma State University - Main Campus, sabiha.parveen@okstate.edu

Neila Donovan
Louisiana State University, ndonovan@lsu.edu

Gina Morris
Oklahoma State University, gina.morris@okstate.edu

See next page for additional authors

DOI: 10.30707/TLCSD7.1.1675490380.822091

Examining the Item-level Psychometric Properties of the Self-Efficacy Questionnaire (SEQ) for Students in Communication Sciences and Disorders

Abstract

The purpose of this study was to examine the psychometric properties of the Self-Efficacy Questionnaire (SEQ), a self-reported questionnaire using Rasch analysis, a one-parameter logistic probability model based on scientific measurement principles. Rasch analysis was completed using the rating scale model. The various analyses allow researchers to examine the item-level psychometric properties of the SEQ, which result in measures that provide evidence for validity, reliability, and sensitivity of the instrument. Rasch analysis demonstrated that the latent trait established by the SEQ, *Perceived Self-efficacy* (called self-efficacy from here on), was a unidimensional construct that could be measured on a linear scale. The instrument demonstrated sound item-level psychometric properties, including a wide span of item difficulty, along with limited ceiling and no floor effects. Person reliability was good, and the SEQ separated raters into at least three statistically different levels of self-efficacy. These results provide evidence for the SEQ's validity, reliability, and sensitivity. Based on this preliminary analysis, the SEQ demonstrated more than adequate item-level psychometric properties for use, although more research needs to be done. Further, instructors could use the SEQ to give preliminary information on whether or not a class assignment leads to increased self-efficacy in undergraduate and graduate students.

Keywords

Perceived self-efficacy, students, Communication Sciences and Disorders, treatment activities, application-based activities

Cover Page Footnote

The current project was partially supported by the Advancing Academic-Research Career Grant from American Speech-Language-Hearing Association awarded to the first author. There are no other relevant financial or nonfinancial disclosures.

Authors

Sabiha Parveen, Neila Donovan, Gina Morris, and Kristi Dickey

Introduction

The scholarship of teaching and learning (SoTL) is related to professional thinking and practicing ways to enhance student learning through peer review and inquiry (Trigwell, 2021). Specific to the field of communication sciences and disorders (CSD), the search for optimal teaching and learning techniques has become more important over the past two decades as costs for higher education have risen and student demands for a more meaningful college education have increased. Some of the existing SoTL research in CSD centers around team-based learning (Epstein, 2016), project-based learning (Burda & Hageman, 2015; Whitehill et al., 2013), and simulations (MacBean et al., 2013; Stead et al., 2020). However, there is a need for additional studies to determine the best practices for the effective teaching and preparation of future speech-language pathologists (SLPs).

While the American Speech-Language-Hearing Association (ASHA) has standards in place for granting individuals graduate degrees in speech-language pathology and professional certification (CCC-SLP), it is up to each CSD program to ensure all individuals applying for certification in speech-language pathology have coursework that includes a proper understanding of assessment and intervention for people with different communication disorders. According to ASHA, these skills may be developed and demonstrated in several ways, including academic coursework, clinical practica, simulations, interprofessional education, labs, examinations, and independent projects (American Speech-Language-Hearing Association, 2022). Although there is evidence that some of these teaching methodologies are effective, there is a need for additional studies in the field of SoTL in CSD, specifically about validating a questionnaire examining the perceived self-efficacy (PSE) of students and practicing SLPs.

Specific to SoTL literature, multiple researchers have reported PSE of early-stage career or practicing SLPs to determine the effectiveness of different academic and clinical training (Beita-Ell & Boyle, 2020; Boyle et al., 2021; Hammer et al., 2003; Parveen & Santhanam, 2021). However, to our knowledge, limited studies have reported the PSE of current CSD students based on academic and/or clinical training (Lee & Schmaman, 1987; Pasupathy & Bogschutz, 2013). Therefore, in the current study, we discuss the development and validity of a new questionnaire that measures the PSE of undergraduate and graduate CSD students based on the completion of an application-based project. The following sections discuss the construct of PSE in more detail, followed by the rationale of the new questionnaire aimed to measure the PSE of undergraduate and graduate CSD students.

Perceived self-efficacy is a person's belief in their capabilities to produce or complete certain tasks (Bandura, 1977). People with high self-efficacy often choose to perform challenging tasks, set higher goals, and work hard to achieve them (Scholz et al., 2002). Researchers in higher education and related fields have reported self-efficacy ratings as an outcome measure of preparation and success among undergraduate and graduate students (Chemers et al., 2001). Further, a positive relationship has been reported between PSE and self-perceived success among SLPs working with specific clients (Beita-Ell & Boyle, 2020). Therefore, application-based projects like the one discussed in the current study may be beneficial in fostering self-motivation and interest in learning more about the field than traditional assignments such as summaries of research articles or observation of therapy sessions. Further, the ability to determine one's ability to replicate or create treatment activities for similar disorders in future practice can also be a source of motivation for students.

Researchers and clinicians have often recognized that the importance or relevance of some variables may not always be seen or evaluated directly. For example, a treatment's specific effect may sometimes be perceived only by the person who received the treatment rather than the clinician administering the treatment. Such traits are latent traits (e.g., pain, effort, or motivation; Baylor et al., 2011; Fries et al., 2007). Likewise, perceived self-efficacy is also a latent trait since it is unobservable. Therefore, gathering a person's perceptions of his/her self-efficacy is necessary. Respondent interviews in qualitative research are one way to collect and analyze respondents' perceptions. However, it may be somewhat cumbersome. A more efficient way to gather respondents' perceptions about a given trait is to use valid and reliable self-report questionnaires such as the one related to PSE.

Patient-reported outcome measures (PRO) typically involve developing large item banks that measure the latent trait. PRO developers use item response theory (IRT; Lord & Novick, 1968) because it results in more precise measurements, which can be used to develop computer-adaptive tests (CATs). The following sections include a brief review of IRT and Rasch analysis, followed by the development of the new SEQ questionnaire.

Self-report Questionnaires. As mentioned above, it is now possible to develop robust self-report questionnaires using more precise methods in developing the questionnaire and demonstrating validity and reliability using the IRT (Bond & Fox, 2007; Lord & Novick, 1968). IRT has also been used in educational measurement for some time now (Bond & Fox, 2007) and has found its way into the field of CSD as researchers and clinicians have begun to understand the benefits of using IRT in test development rather than the Classical Test Theory (Baylor et al., 2011; Donovan, 2018).

One of the main benefits is that using the IRT methodology results in measurement precision by transforming ordinal data into interval data. Person-ability and item difficulty are calibrated on a single interval scale which can be used in ways ordinal data cannot, such as to perform inferential analysis of results across groups or between individuals at different time points. In addition, because the items are ordered by difficulty, a researcher or clinician can determine what tasks a person finds beneficial to learning over another task. Alternatively, as respondents are ordered by their ability level, they can be identified as those with low PSE and those with high PSE or struggling with certain tasks but not others. Interested readers are encouraged to an online resource that has a complete bibliography on PRO development and IRT methodology (National Institute of Health Patient-Reported Outcomes Measurement Information System, n.d.). A tutorial written for clinicians and researchers in CSD is also available (Baylor et al., 2011).

Rasch analysis (Rasch, 1960/1980) is a mathematical model based upon a latent trait (e.g., self-efficacy), which accomplishes probabilistic conjoint additivity (conjoint means measurement of persons and items on the same scale, and additivity is the equal-interval property of the scale (Granger, 2008; Wright, 1999). The resulting interval scale is calibrated for both *self-rated ability* and *item difficulty* for the latent trait (e.g., for this report, self-efficacy). The Rasch formula is based on the same formula as the one-parameter (1-PL) IRT model developed in the 1950s and

1960s to meet the needs of large-scale educational testing (Bock, 1997). However, Rasch sought a model that adhered to the principles of objective measurement (one of the standards of scientific measurement) defined as:

> "Objective measurement is the repetition of a unit amount that maintains its size, within an allowable range of error, no matter which instrument, intended to measure the variable of interest, is used and no matter who or what relevant person or thing is measured. (Program Committee of the Institute for Objective Measurement, 2000)."

To meet the requirements of objective measurement, a questionnaire must measure one latent trait (e.g., a ruler measures length). In addition, the instrument must have equal and additive units that do not change (e.g., the inch or the centimeter is the basic unit of measure on a ruler). Finally, the instrument may have various forms but must measure anything for the trait of interest (e.g., length may be measured with a ruler, a meter stick, or a tape measure). Some researchers have dismissed the value of self-report questionnaires because, in the past, some were poorly developed and/or contained inherent biases that could lead to erroneous conclusions. Developing a questionnaire using the Rasch model alleviates some of those concerns.

The intervality of Rasch scaling gives clinicians and researchers additional benefits, such as comparing one individual's scores at different time points (e.g., pre-post treatment outcome testing) and feeling confident about the meaning of the change. Additionally, researchers could compare group data and perform inferential statistical testing using interval data. Finally, the Rasch analysis provides information about a questionnaire's quality, comparable to the CTT measures of validity, reliability, and sensitivity. With these benefits in mind, the investigators in the current study chose Rasch analysis to determine the item-level psychometric properties of the SEQ.

Development of the Self-Efficacy Questionnaire (SEQ). Fink (2003) has extensively discussed the major educational goals and significant learning based on six dimensions- learning how to learn, caring, human dimension, integration, application, and foundational knowledge. The treatment kit project was developed based on the principle of application that facilitates critical thinking, practical thinking, creativity, and the management of complex projects. The treatment kit project provided opportunities to improve the curriculum and overall learning of the students. These benefits became obvious through formal course evaluations and correspondence from recent alumni. Each year the students commented on the experience of making activities by themselves and getting opportunities to directly apply some of the learned concepts to their therapy activities. As a result, they reported feeling more prepared and confident about using these activities with their current and future clients.

The first author then searched the literature to identify any existing self-efficacy questionnaire currently used in CSD that can formally help assess the effectiveness of the new application-based project. However, no questionnaires were found specific to students' self-efficacy based on learning from application-based projects. Therefore, the authors compiled a set of questions based on existing literature (Bandura, 1997; Bandura, 2006; Chemers et al., 2001; Lee & Schmaman, 2001; Nwosu & Okoye, 2014; Plotts, 2017). These questions were then shared with three scholars with experience in SOTL research for clarity and internal validity. Finally, the first, third, and fourth authors re-reviewed the questionnaire for clarity and finalized the questionnaire with a set of 13 Likert-type questions.

Purpose. The current study used Rasch Analysis to investigate the item-level psychometric properties of the Self-Efficacy Questionnaire (SEQ). The current study includes findings from nine out of 13 Likert-type questions designed to measure students' self-efficacy. To achieve this purpose, the investigators asked the following experimental questions:

1. Does the SEQ represent a unidimensional construct?
2. Does the SEQ demonstrate sound item-level psychometric properties that demonstrate self-efficacy can be measured on a linear scale?
3. Does the SEQ demonstrate an adequate range of a person's self-rated ability and item difficulty to separate respondents into at least two levels of ability?

Methods

Design. This was a prospective study to investigate the item-level psychometric properties of the SEQ. The Human Subjects Review Board approved the study before data collection. All participants completed an Informed Consent before participating in any screening or data collection procedures. Mathematicians and survey developers debate the adequate sample size for IRT and Rasch analysis. Proponents of IRT, which may use more complex models and several parameters rather than a single parameter, advocate very large sample sizes. However, for a 1-PL Rasch analysis, authors have suggested that a well-target sample size of 30 was sufficient to obtain item calibration stability on a polytomous scale within ±1 logit with 99% confidence (Linacre, 1994). Therefore, the initial analysis used ratings from 77 respondents.

Participants. A total of 79 students, including 55 undergraduate and 24 graduate students enrolled in a CSD program, completed the SEQ during fall 2019 in two courses, *Speech and Language Development* and *TBI & Dementia*, respectively. Both courses were taught by the first author, who was blinded to student participation during the data collection period. Two respondents who demonstrated extremely positive scores were removed from the initial analysis, leaving 77 participants. Inclusion criteria included participants 18 years or older enrolled in one of the two courses mentioned above. There were no other specific exclusion criteria. No other demographic information (including age, gender, and ethnicity) was collected for the study.

Procedures. The SEQ consisted of nine items with two 5-unit rating scales. The first eight questions had a Likert rating ranging from 1 = definitely cannot do it to 5 = definitely can do it. The final item had a Likert rating ranging from 1 = not at all to 5 = extremely. Data coding before Rasch analysis allowed the two rating scales to be incorporated into a single 1 to 5 scale. The SEQ is attached as Appendix A. The third and fourth authors administered the survey twice during the fall 2019 semester (once at the beginning and once at the end). All surveys were completed in person. The third and fourth authors deidentified the collected survey responses following data collection. Thus, only the course information was retained on the surveys. Once the surveys were collected, research assistants (blinded to participants' identity) entered item responses in the database using a double-entry system to ensure accuracy. Again, blinded to participants' identity, the second author prepared the Rasch analysis control files and performed all subsequent Rasch analyses.

Data Analysis. The authors chose to use WINSTEPS (v. 5.1.4.0) to determine the item-level psychometric properties of the SEQ. The data were polytomous. Therefore, the rating scale model was selected. For readers interested in the specific mathematical formula for the Rasch rating scale model, please refer to a chapter by Andersen (1997). In addition, the following aspects important to measurement quality were analyzed.

Unidimensionality. Rasch modeling requires meeting three assumptions: a) the latent trait must be unidimensional; b) the items must differentiate people with different ability levels, and c) the items must be locally independent. A Standardized Residuals Principal Components Analysis (PCAR) was used to determine the SEC's dimensionality based on the following criteria: a) the first contrast must have an eigenvalue < 2 and disattenuated standardized residual correlations (error has been removed from analysis) > 0.70; b) the largest standardized correlations are < 0.70; and c) person separation must be > 2 (Linacre, 2018, 2021; Raiche, 2005). The person separation index measures the number of statistically distinct ability levels into which respondents can be separated. A person separation index > 2 (standard error units) is needed to attain a .8 reliability coefficient (Bond & Fox, 2007). The person separation index > 2 results in at least three statistically different levels of ability.

***Item fit statistics*.** Item fit statistics provide information about the variation between the model and the actual data. For example, an item fit statistic of 1.0 indicates that the actual responses match the modeled response probability perfectly—the closer to 1.0, the better the item fit. For statisticians trained in CTT, the item fit statistics relate to chi-squares based on standardized item/person residuals (Wright & Linacre, 1994). A developer may choose from a range of acceptable item fit criteria depending on the purpose of the analysis (Linacre, 2003). For example, if a researcher has a large data set and is seeking tight precision, he/she may set narrow item-fit criteria. On the other hand, setting wider item-fit criteria is acceptable for a new questionnaire such as the SEQ (Linacre, 2021). Therefore, the authors established acceptable item fit criteria < .05 and ≥ 1.5 mean square (MnSq), indicating that the fit is productive for measurement. Misfitting items may be removed until all items meet acceptable item fit criteria. Or, if a poorly fitting item is deemed critical to the latent theoretical construct, it may be retained (Bond & Fox, 2007).

***Item difficulty hierarchy*.** The Person/Item Map (Boone, Staver, and Yale, 2014) provides a visual interpretation of the Rasch analysis showing the item difficulty hierarchy and the person's ability calibrated along a single interval logit scale. As an inch is the unit of measure on a ruler, the logit is the unit of measure in Rasch analysis. A logit is "the distance along the line of the variable that increases the odds of observing the event specified in the measurement model by a factor of 2.718, the value of "e" the base of "natural" or Napierian logarithms used for the calculation of "log-" odds" (Linacre & Wright, 1989, p. 54). An acceptable range of item difficulty ranges between -2 to +2 logits (Bond & Fox, 2007). A test or questionnaire will tap a full range of respondents' abilities in the ideal. In essence, there should be no ceiling or floor effects. However, many tests and questionnaires developed using CTT have ceiling and floor effects because the full range of ability and item difficulty is not investigated. A well-developed instrument should have no more than 10% ceiling and floor effects (Bond & Fox, 2007).

Item mean-person means. Comparing the item difficulty and self-rated ability mean indicates how well the items have captured the person's self-efficacy (i.e., the main latent trait). The closer the means are to 0 logits, the better. A perfect match occurs where both means are equal to 0. An acceptable item mean-person means the match is 0 ± 1 logit (Bond & Fox, 2007). A large mismatch between item mean and person means suggests that the items are either too easy or too difficult for the respondents depending on the direction of the mismatch.

Person reliability. The construct of person reliability is comparable to Cronbach's alpha (Cronbach, 1962), a measure of internal consistency ranging from 0 to 1. Higher values indicate more reliable response measures. The rule of thumb suggests reliability in the following ranges 0.7 to .79 is acceptable, .8 to .89 is good, and 0.9 and above is excellent (Cronbach, 1962). Because the SEQ was a new instrument with unknown psychometric properties, we established acceptable person reliability at .70 or above.

Results

As mentioned above, one benefit of Rasch analysis is that it allows questionnaire developers to refine the data to improve an instrument's psychometric properties. For example, a misfitting item indicates unnecessary or unexpected "noise" in the model and can be removed. On the other hand, the theoretical foundation of the questionnaire may require that even a misfitting item remains in the analysis (Bond & Fox, 2007). Likewise, respondents who provide unexpected or aberrant responses may also be removed. Although the original data set included 79 respondents, 77 were included in the initial analysis. After the first analysis, eight of the thirteen items fit the model and one misfit but were important to the theoretical construct of self-efficacy, so it was retained for further analyses. However, 12 respondents demonstrated significant misfits and were removed at this time. Appendix B includes the data for misfitting person data. A second analysis was then performed using 65 respondents and nine items. The results of this analysis were used to answer the three research questions.

The study's first research question aimed to examine whether or not the SEQ data represented a unidimensional construct. Results indicated that the SEQ demonstrated a unidimensional construct based on results that met the earlier three assumptions. First, the Rasch PCAR indicated that the variance explained by the Rasch dimension (self-efficacy) was 60.1%. Second, the largest secondary dimension (i.e., the first contrast in the standardized residuals) accounted for 8.3% (1.87 eigenvalue) of unexplained variance, and the disattenuated correlations (error removed) between the three clusters of items were highly correlated, again indicating that items were comprised of a single contrast. See Table 1 for correlations among the items' three clusters of standardized residual correlations.

The SEQ items also met the assumption for local independence based on examining the correlations among standardized residuals. None of the correlations rose to the level of 0.7 and therefore were not concerned with local dependency. Table 2 summarizes the largest standardized residual correlations to identify the dependent items.

DOI: 10.30707/TLCSD7.1.1675490380.822091

Table 1

Standardized Residual Correlations Among the Item Cluster Contrasts

PCA Contrast	Item Clusters	Pearson Correlation	Disattenuated Correlation
1	1-3	0.447	0.816
1	1-2	0.636	1.000
1	2-3	0.410	0.768

Table 2

Largest Standardized Residual Correlations Used to Identify Dependent Items

Correlations	Entry #	Entry #
0.25	6 explTVidea	8 explarticle
0.16	4 oldhelpnew	5 recalltest
0.12	1 figureout	2 canlearn
-0.29	8 explarticle	9 futurekits
-0.44	3 setgoals	6 explTVidea
-0.41	5 recalltest	7 explconftopic
-0.40	6 explTVidea	9 futurekits
-0.28	2 canlearn	5 recalltest
-0.28	1 figureout	5 recalltest
-0.26	4 oldhelpnew	8 explarticle
-0.26	4 oldhelpnew	9 futurekits
-0.25	1 figureout	8 explarticle
-0.25	2 canlearn	4 oldhelpnew
-0.24	5 recalltest	8 explarticle
-0.23	2 canlearn	7 explconftopic
-0.23	3 setgoals	7 explconftopic
-0.17	2 canlearn	6 explTVidea
-0.16	1 figureout	4 oldhelpnew
-0.14	1 figureout	7 explconftopic
-0.13	5 recalltest	9 futurekits

Item fit. Rasch's analysis results indicated the following: Using the established criteria (MnSq < .5 ≥ 1.5), the SEQ had one misfitting item, "recall information for a test" (Infit MnSq 1.61). The item was retained in the analysis because an instructor's self-efficacy during test-taking was considered important information. It is permissible to retain a misfitting item if important to the construct (Bond & Fox, 2007).

Item difficulty hierarchy. The SEQ9 item hierarchy spanned 4.68 logits (-2.20 to 2.49). This is above the established criteria (-2 to +2 logit range). Only 4.6% of respondents (three) demonstrated a ceiling effect, and no floor effect was observed. The analysis showed the item that required the least self-efficacy to endorse (I can learn) and the item that required the most self-efficacy to endorse (I can explain a conference presentation topic), as was theorized by the developers. See Table 3 for the summary of the item level difficulty calibrations for the SEQ.

Table 3

SEQ Summary of Item Level Difficulty Calibrations for the SEQ

	Item	Total Score	Total Count	Item Measure	Model S.E.	Infit MnSQ
7	explain conference topic	226	65	2.48	0.22	0.85
8	explain research article	228	65	2.38	0.22	0.60
6	explain T.V. show	247	65	1.42	0.23	0.86
1	figure things out	277	65	-0.28	0.25	0.89
4	use old info to learn new	278	65	-0.35	0.25	0.87
5	recall info for test	279	65	-0.41	0.25	1.61
9	future kits	287	65	-0.93	0.26	1.27
3	set goals	303	65	-2.11	0.29	0.92
2	can learn	304	65	-2.20	0.30	1.13

Note. S.E. = standard error; MNSQ = mean square

The third research question was whether or not the SEQ demonstrated an adequate self-efficacy range and item difficulty separating respondents into > 2 levels. Results indicated that the SEQ separated SELF-efficacy and item difficulty into > 2 levels. Therefore, the question was answered based on the following results.

Item mean-person mean. The SEQ item means (0 logits) and person mean (2.32 logits) did not quite fall within the ideal range of 0 ± 1 logit. However, a .32 logit difference was deemed acceptable for the initial analysis because it will inform further SEQ development (discussed in the next section).

Person reliability. The respondents' responses demonstrated good reliability (0.82) for rating self-efficacy. This statistic is comparable to Cronbach's alpha and indicates good internal consistency.

Person separation. The person separation index of 2.14 exceeded the criteria of > 2. This suggested that the SEQ was sensitive to at least three statistically different levels of self-efficacy.

Table 4 summarizes the established criteria compared to the results of the SEQ Rasch analysis. In addition to answering the question with the numeric results described above, the Rasch analysis provides a visual map illustrating the results along the logit scale derived for self-efficacy and item difficulty on the SEQ. The benefits of the Map will be discussed in the next section.

https://ir.library.illinoisstate.edu/tlcsd/vol7/iss1/3
DOI: 10.30707/TLCSD7.1.1675490380.822091

Table 4

Summary of Rasch Analysis Criteria for Sound Item-Level Psychometrics Compared to Established Criteria

Rasch Analysis	Established Criteria	Results
Item in-fit statistics	MnSq <.5 to >1.5	*Retain 9 Items
Item difficulty hierarchy	-2 to + 2 logit span	4.68 logit span
Ceiling and floor effects	<10%	4.6% Ceiling, 0% Floor
Match Item Mean-Person Mean	0 ± 1 logit	Item M = 0; Person M = 2.32
Person Reliability	0.7 - 0.79 acceptable	0.82 good
Person Separation Index	≥ 2	2.14
Item Reliability	0.7 - 0.79 acceptable	.97 excellent
Item Separation	> 2	6.06

Note: Item #5 "recall material for a test (1.61 MnSq) was retained in the SEQ because it was important to the theoretical foundation of the questionnaire (Bond & Fox, 2007).

Figure 1 includes a map of students' self-efficacy ratings and SEQ item difficulty hierarchy based on Rasch analysis. This figure provides a visual illustration of the SEQs item level psychometric properties. The orientation box explains the map arrangement and notations. The shaded area relates to information about students' self-efficacy; the non-shaded area relates to the item difficulty hierarchy (which items are easier or harder for students with different amounts of self efficacy.

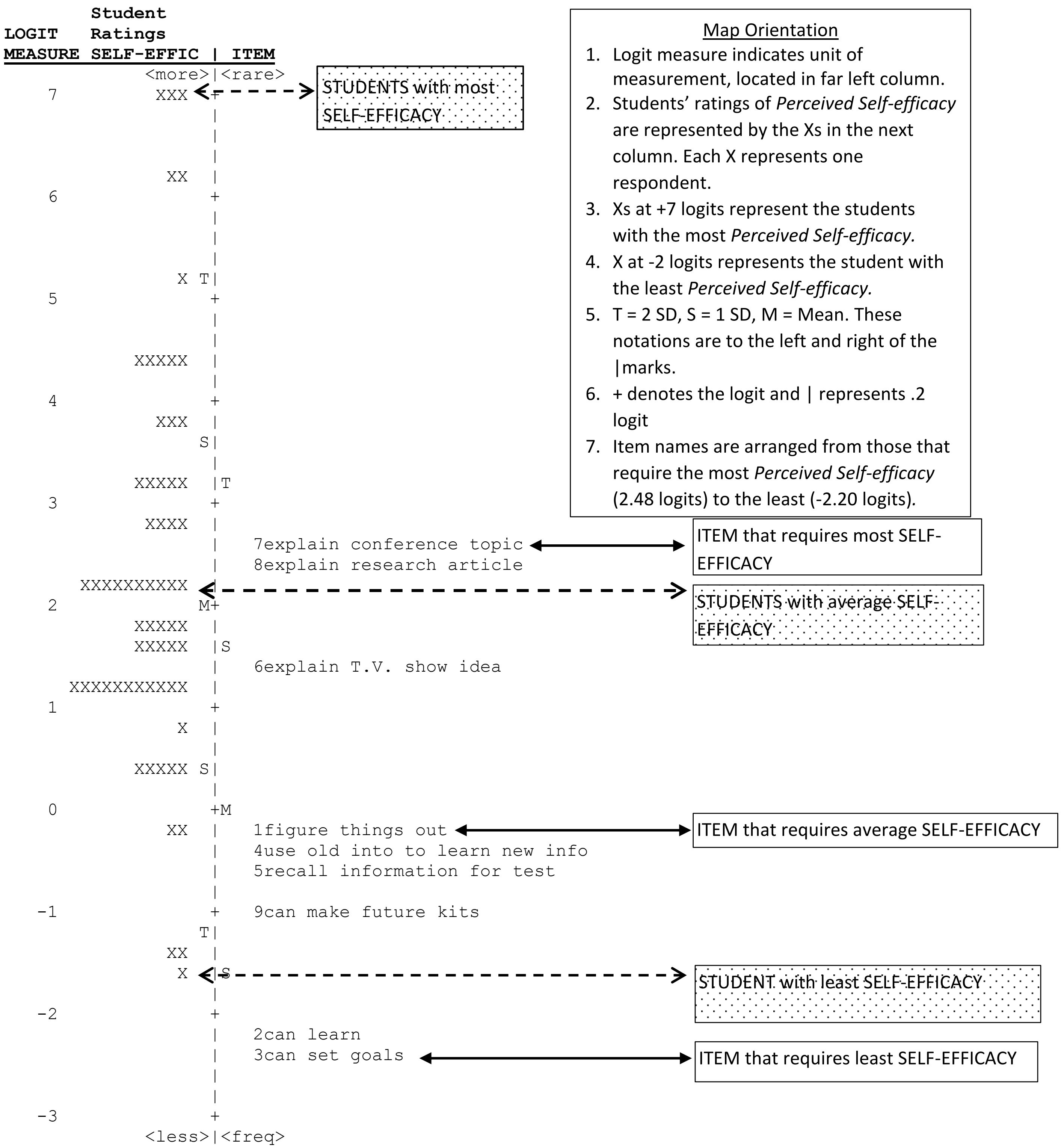
Discussion

This study aimed to investigate the item-level psychometric properties of the SEQ, an instrument designed to measure the self-efficacy of CSD students based on the completion of treatment activities for specific disorders. Results indicated that the SEQ questionnaire met the assumptions of the Rasch analysis as it accounted for 61% of the variance in the model. In addition, the person's self-reliability score indicated good internal consistency. Further, the nine questions on the SEQ that were analyzed seemed to capture the wide range of abilities of the students.

Providing CSD students with assignments that help them apply what is learned in the classroom and increase their feelings of self-efficacy—"I can do it"—is an area that needs further study. Researchers and teachers must have good ways to determine their assignments/pedagogy outcomes with any newly emerging construct. Well-designed self-report questionnaires may be an ideal way to measure pre- and post-assignment changes in self-efficacy since, as a latent trait, it is not readily observable. The SEQ was developed to address this need. Rasch's analysis indicated that the SEQ demonstrated sound item-level psychometric properties like measures of internal consistency used in CTT. The following sections include a brief discussion of how the analysis results could be used, some limitations of the study, and future research directions.

Figure 1

Map of Students' Self-Efficacy Ratings and SEQ Item Difficulty Hierarchy Based on Rasch Analysis.



As discussed earlier, Rasch (1960/1980) described the principles of objective measurement upon which the current research questions were based. First, objective measurement requires that the instrument measure a single trait (e.g., a ruler measures length). In this report, as the investigators operationalized, students' perceived self-efficacy was a unidimensional latent trait. This is an interesting finding, given that many ways to define and measure self-efficacy exist.

Next, objective measurement principles require an instrument to have equal, additive, and invariant units of measurement (e.g., the inch is the basic unit of measure on a ruler). Rasch analysis transforms ordinal data into interval scores for both item difficulty and a person's self-rated ability along a single interval scale measured in logits from, in this case, students with the least self-efficacy to those with the most. Items are also ordered from those easy for a person with less self-efficacy to endorse items requiring more self-efficacy, referred to as the item difficulty hierarchy. Results indicated that even students with the lowest SEQ ratings could endorse "I can learn what is being taught in class this semester." Still, only students with higher self-efficacy could endorse the item." Currently, I am confident that I can explain the main ideas of a professional conference presentation dealing with course topics to another person."

Furthermore, the range of item difficulty measures and the person's self-rated ability measures were more than adequate, with logit spans well above the requisite -2 to + 2 logits, demonstrating that the SEQ had captured three statistically significant levels of self-efficacy in only nine questions. The Map shown in Figure 1 showed gaps among the items, where more questions could be added in the future if the investigators felt it was warranted. Likewise, considering that the match between Person and Item Means was adequate but not optimal, the gaps between items on the Map indicate where more difficult items could be added to tap the higher self-efficacy ratings of the students. For example, perhaps items could be added to measure self-efficacy in leading class discussions, making class presentations, presenting research studies or case studies to the class, or leading a grand rounds presentation in a clinical setting.

As discussed, there are two main benefits of interval scaling. First, it allows researchers to perform statistical inference testing, which should not be done with ordinal or nominal data. Second, and perhaps more importantly, interval scaling provides clinicians with numbers that carry meaning, unlike reporting percentages or raw scores, which do not provide information on the person's self-rated ability since one can get different questions right and wrong and still have a certain percent score. However, if a person progresses from +5 to +10 on any interval scale, they have doubled their score, which would be deemed a solid indicator of progress. In addition, teachers and investigators could compare SEQ pre- and post-learning group or individual scores to demonstrate improved self-efficacy or determine where more instruction is needed. For example, in the Figure 1 Map, respondents are represented by Xs. However, an instructor could use a Map that shows students ranked by I.D. information (not presented here due to space limitations) to identify students with low self-efficacy early in the semester.

The self-efficacy questionnaire can be used either by itself or with other assessment metrics, such as class presentations and written reports to determine the effectiveness of different application projects. Therefore, future studies can examine multiple datapoints from the application-based classroom projects, including the objective ratings on the SEQ questionnaire, the qualitative analyses of the students' perceptions, and end-of-semester performance. Positive relationships between the different metrics are likely to indicate that the students' self-reported efficacy and performance on the project are linked. Still, the supposition has not been tested so far.

Study Limitations and Future Directions

The current study presents the preliminary results of the SEQ Rasch analysis based on a relatively small sample of undergraduate and graduate SLP students. As mentioned earlier, there is an ongoing debate about the appropriate sample size for Rasch analysis and other IRT models. However, the literature provides evidence that a small well-grouped sample may be adequate for initial analysis (Linacre, 1994), although item calibration may be less stable (Chen et al., 2013). Therefore, a study with more respondents will be an important next step in SEQ development. Furthermore, 20 items appear important to item calibration (Chengappa & Kumar, 2008). Therefore, the SEQ could be further refined by adding additional items that would measure self-efficacy more broadly, particularly by creating items that tap more challenging activities to one's self-confidence.

One way to understand students' self-efficacy would be to conduct a qualitative research study. Students are asked to review the SEQ items and provide input on what more could be asked or ask students how they define self-efficacy. Further, a mixed-methods design of both quantitative and qualitative responses can help identify the specific factors that influence or shape the PSE of students training to be SLPs. Therefore, future research will focus on detailed comparisons between graduate and undergraduate student cohorts across multiple semesters to determine if there are any group differences and determine factors that contribute most to the growth and trajectory of self-efficacy. Finally, additional studies will report the relationships between different data points for application-based projects, including questionnaire ratings, mid-term drafts, final presentations, and related reports.

Conclusion

The current study discusses the psychometric properties of a new self-efficacy questionnaire (SEQ) for CSD students based on application-based treatment activities. Results indicated strong psychometric properties of the SEQ, thereby indicating its' relevance and utility. Future studies will focus on findings of the SEQ questionnaire based on larger cohorts and across multiple semesters.

Author Disclosures

The current project was partially supported by the Advancing Academic-Research Career Grant from American Speech-Language-Hearing Association awarded to the first author. There are no other relevant financial or nonfinancial disclosures.

References

American Speech-Language-Hearing Association (2022). 2020 Standards and Implementation Procedures for the Certificate of Clinical Competence in Speech-Language Pathology. https://www.asha.org/certification/2020-slp-certification-standards/

Andersen, E. B. (1997). The Rating Scale Model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*. New York, NY: Springer.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy Beliefs of Adolescents*, 307-337.

Baylor, C., Hula, W., Donovan, N. J., Doyle, P. J., Kendall, D., & Yorkston, K. (2011). An introduction to item response theory and Rasch models for speech-language pathologists. *American Journal of Speech-Language Pathology, 20,* 243-259. https://doi.org/10.1044/1058-0360(2011/10-0079)

Beita-Ell, C., & Boyle, M.P. (2020). School-based speech-language pathologists' perceived self-efficacy in conducting multidimensional treatment with children who stutter. *Language, Speech, and Hearing Services in Schools, 51*, 1172-1186. https://doi.org/10.1044/2020_LSHSS-20-00044

Boyle, M.P., Beita-Ell, C., & Chagachbanian, N.J. (2021). Perceptions of self-efficacy in providing multidimensional school-age stuttering therapy among board certified fluency specialists in the United States. *Journal of Fluency Disorders, 69*, 105862. https://www.sciencedirect.com/science/article/abs/pii/S0094730X21000413?via%3Dihub

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, *16*, 21–33.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. New York: Springer.

Burda, A.N., & Hageman, C.F. (2015). Problem-based learning in speech-language pathology: Format and feedback. *Contemporary Issues in Communication Sciences and Disorders, 42*, 47-71. https://pubs.asha.org/doi/pdf/10.1044/cicsd_42_S_47

Bureau of Labor Statistics, U.S. Department of Labor. (2022, September 28). *Occupational Outlook Handbook*. Speech-Language Pathologists https://www.bls.gov/ooh/healthcare/speech-language-pathologists.htm

Chemers, M.M., Hu, L., & Garcia, B.F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology, 93*(1), 55-64. https://doi.org/10.1037/0022-0663.93.1.55

Chen, W. H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2013). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research, 23*, 485-493. https://doi.org/10.1007/s11136-013-0487-5

Chengappa, S. K., & Kumar, R. (2008). Normative and Clinical Data on the Kannada Version of Western Aphasia Battery (WAB-K). *Language in India: Strength for Today and Bright Hope for Tomorrow, 8*, 1–15.

Cronbach, L. J. (1962). Coefficient alpha and the internal structure of a test. *Psychometrika, 16*, 297-334.

Donovan, N. J. (2018). Examining the item-level psychometric properties of the Communication Effectiveness Survey-Revised for people with Parkinson's disease and dysarthria. *Clinical Archives in Communication Disorders, 3*(1), 42-51. http://dx.doi.org/10.21849/cacd.2018.00269

Epstein, B. (2016). Five heads are better than one: Preliminary results of team-based learning in a communication disorders graduate course. *International Journal of Language and Communication Disorders, 51*, 44-60. https://doi.org/10.1111/1460-6984.12184

Fink, D.L. (2003). *Creating significant learning experiences: An integrated approach to designing college courses*. San Franscisco: Jossey-Bass.

Fries, J. F., Bruce, B., Bjorner, J. B., & Rose, M. (2006). More relevant, precise, and efficient items for assessment of physical function and disability: Moving beyond the classic instruments. *Annals of the Rheumatic Diseases, 65*, 16-21. https://doi.org/10.1136/ard.2006.059279

Granger. (2008). Rasch Analysis is Important to Understand and Use for Measurement. *Rasch Measurement Transactions, 21*(3), 1122-1123.

Hammer, C.S., Detwiler, J.S., Detwiler, J., Blood, G.W., & Qualls, C.D. (2004). Speech-language pathologists' training and confidence in serving Spanish-English Bilingual children. *Journal of Communication Disorders, 37*, 91-108. https://doi.org/10.1016/j.jcomdis.2003.07.002

Lee, C., & Schmaman, F. (1987). Self-efficacy as a predictor of clinical skills among speech pathology students. *Higher Education, 16*, 407-416.

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

Linacre, J. M. (2003). Size vs. significance: Standardized chi-square fit statistic. *Rasch Measurement Transactions, 17*, 918.

Linacre, J. M. (2021). Winsteps® (v5.1.4.0). Rasch measurement computer program and User's Guide. Beaverton, Oregon: Winsteps.com.Linacre, J.M.

Linacre, J. M., & Wright, B. D., (1989). The "Length" of a Logit. *Rasch Measurement Transactions, 3*(2), 54-55.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

MacBean, N., Theodoros, D., Davidson, B., & Hill, A.E. (2015). Simulated learning environments in speech-language pathology: An Australian response. *International Journal of Speech-Language Pathology, 15*(3), 345-357.

National Institute of Health Patient-Reported Outcomes Measurement Information System (n.d.) http://www.healthmeasures.net/explore-measurement-systems/promis

Nwosu, K.C., & Okoye, R.O. (2014). Students' self-efficacy and self-rating scores as predictors of their academic achievement. *Journal of Educational and Social Research, 4*(3), 223-228.

Parveen, S., & Santhanam, S. (2021). Speech-language pathologists' perceived competence in working with culturally and linguistically diverse clients in the United States. *Communication Disorders Quarterly, 42*(3), 166-176. https://doi.org/10.1177/1525740120915205

Pasupathy, R., & Bogschutz, R.J. (2013). An investigation of graduate speech-language pathology students' SLP clinical self-efficacy. *Contemporary Issues in Communication Sciences and Disorders*, *40*, 151-159. https://pubs.asha.org/doi/pdf/10.1044/cicsd_40_F_151

Plotts, K. (2017). *An inquiry into the development of future speech-language pathologists: A mixed methods study* [Doctoral dissertation, Rowan University]. Theses and Dissertations. https://www.proquest.com/docview/1964804254/fulltextPDF/9DBA5D75A023466EPQ/1?accountid=4117

Program Committee of the Institute for Objective Measurement (2000, December). *Definition of Objective Measurement*. https://www.rasch.org/define.htm

Raiche, G. (2005). Critical eigenvalue sizes in standardized residual principal components analysis. *Rasch Measurement Transactions*, *19*(1), 1012.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Scholz, U., Dona, B.G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? *European Journal of Psychological Assessment*, *18*(3), 242-251.

Stead, A., Lemoncello, R., Fitzgerald, C., Fryer, M., Frost, M., & Palmer, R. (2020). Clinical simulations in academic courses: Four case studies across the medical SLP graduate curriculum. *Teaching and Learning in Communication Sciences and Disorders*, *4*(3), Article 6. https://doi.org/10.30707/TLCSD4.3/ACVJ1784

Trigwell, K. (2017). Scholarship of teaching and learning. In L. Hunt and D. Chalmers (Eds.), *University Teaching in Focus: A learning-centered approach*. Routledge.

Whitehill, T.L., Bridges, S., & Chan, K. (2013). Problem-based learning (PBL) and speech-language pathology: A tutorial. *Clinical Linguistics & Phonetics*, *28* (1-2), 5-23. https://doi.org/10.3109/02699206.2013.821524

Wright, B. D., & Linacre, J. M. (1994). Reasonable item mean-square fit value. *Rasch Measurement Transactions*, *8*, 370.

Wright, B.D. (1999). Fundamental measurement for psychology. In S.E. Embretson & S.L. Hershberger (Eds.) *The new rules of measurement: What every psychologist and educator should know*. Erlbaum: Mahway, NJ.

Appendix A
Self-Efficacy Questionnaire (SEQ)

Section 1: Self-Reported Competency

1. I can figure out anything related to this course if I try hard enough.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

2. I can learn what is being taught in class this semester.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

3. I am confident that I will achieve the goals that I set for myself.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

4. When I am trying to understand a new topic, I can associate new concepts with old ones sufficiently well to remember them.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

5. When I have trouble recalling an abstract or difficult concept, I can think of a good example that will help me remember it.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

6. Currently, I am confident that I can explain the main ideas of a television documentary dealing with course topics to another person.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

7. Currently, I am confident that I can explain the main ideas of a professional conference presentation dealing with course topics to another person.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

8. Currently, I am confident that I can explain the main ideas of a research article dealing with course topics to another person.				
Definitely	*Probably*	*Maybe*	*Probably*	*Definitely*
Cannot Do it	*Cannot*	*Can*	*Can*	*Can Do it*
1	2	3	4	5

DOI: 10.30707/TLCSD7.1.1675490380.822091

Section II: Impact of Treatment Kits

Please select the statement that seems appropriate to your experiences.

◯ I have created and used treatment kits before. (if so, list the course and semester):______________________________

◯ I have never created treatment kits and this is my first experience.

9. Based on my experiences to date, I believe that creating a treatment kit *will increase* or *has increased* my ability to create future treatment kits for other disorders.

Not at all	*Slightly*	*Somewhat*	*Moderately*	*Extremely*
1	2	3	4	5

10. Based on my experiences to date, I believe that making a treatment kit *will increase* or *has increased* my ability to perform a successful treatment session.

Not at all	*Slightly*	*Somewhat*	*Moderately*	*Extremely*
1	2	3	4	5

11. Based on my experiences to date, I believe that making a treatment kit *will increase* or *has increased* my ability to mentor another student in creating treatment kits.

Not at all	*Slightly*	*Somewhat*	*Moderately*	*Extremely*
1	2	3	4	5

12. Based on my experiences to date, I believe that making a treatment kit *will increase* or *has increased* my knowledge related to interventions for specific disorders.

Not at all	*Slightly*	*Somewhat*	*Moderately*	*Extremely*
1	2	3	4	5

13. Based on my experiences to date, I believe that making a treatment kit *will increase* or *has increased* my ability to share the rationale and effectiveness of intervention activities with clients and their families in a professional and effective manner.

Not at all	*Slightly*	*Somewhat*	*Moderately*	*Extremely*
1	2	3	4	5

Appendix B

Misfitting persons from the SEQ that were removed from Rasch analysis

Person ID	Infit MnSq (> 1.5)
U19019	1.61
U19017	1.82
U19055	3.27
U19020	2.22
U19037	2.26
U19049	2.00
U19008	1.79
M19021	1.85
U19006	1.70
U19047	1.70
U19058	1.62

Note: U = Undergraduate, M = Masters

https://ir.library.illinoisstate.edu/tlcsd/vol7/iss1/3
DOI: 10.30707/TLCSD7.1.1675490380.822091