

Rasch analysis of the Self-Efficacy for Therapeutic Use of Self questionnaire in Norwegian occupational therapy students

Chia-Wei Fan^a, Farzaneh Yazdani^b, Tove Carstensen^c and Tore Bonsaksen^{d,e}

^aDepartment of Occupational Therapy, AdventHealth University, Orlando, FL, USA; ^bFaculty of Health and Life Sciences, Oxford Brookes University, Oxford, UK; ^cDepartment of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway; ^dDepartment of Occupational Therapy, Prosthetics and Orthotics, Faculty of Health Sciences, OsloMet – Oslo Metropolitan University, Oslo, Norway; ^eFaculty of Health Studies, VID Specialized University, Sandnes, Norway

ABSTRACT

Background: Research suggested combining modern test theory with classical test theory to achieve comprehensive evaluation of an assessment tool. However, the Norwegian Self-Efficacy for Therapeutic Use of Self questionnaire has not yet been examined by the modern test theory.

Aims/objective: This study aims to examine psychometric properties of the Norwegian Self-Efficacy for Therapeutic Use of Self questionnaire by using Rasch analysis.

Material and methods: One hundred and eleven occupational therapy students from two universities in Norway completed the questionnaire across four time points. Rasch analysis was used to examine the appropriateness of the rating scales, unidimensionality, person response validity, item/person separation, and the Wright map.

Results: The ten-point rating scale did not fully maximise the measurement potentials. Unidimensionality was confirmed except for two items. Person response validity needs further investigation. Excellent person/item separation and Wright map were found.

Conclusion: This study supports the psychometric properties of the Norwegian Self-Efficacy for Therapeutic Use of Self questionnaire in assessing self-efficacy in therapeutic encounters. Further research is needed to address the misfit items and the rating scale issue.

Significance: Rasch analysis showed that the Norwegian Self-Efficacy for Therapeutic Use of Self questionnaire is promising to be used as a reliable and valid tool.

Introduction

Self-efficacy refers to an individual's self-perceptions of effectiveness in using one's own ability to achieve desired outcomes from one's actions [1]. A certain level of self-efficacy is essential in daily living to manage routines and complete daily tasks. Health professionals importantly need self-efficacy in their ability to handle interactions with their clients. To be a competent therapist, one needs not only to be prepared with adequate professional knowledge, but also be equipped with the appropriate skills to convey one's thoughts and reason with clients [2]. Having appropriate self-efficacy for therapeutic use of self can promote a good interpersonal relationship, and ultimately lead to higher client satisfaction and better therapeutic compliance [3,4]. Occupational therapy educators are challenged to educate students about the strategies of therapeutic use of self and to help students build self-efficacy to apply these strategies

comfortably. Recent studies in Norway investigated occupational therapy students' self-efficacy for therapeutic use of self, as conceptualised within the Intentional Relationship Model (IRM) [4]. The results revealed that the students experienced increasing self-efficacy in a 3-month [5], 10-month [6], and 16-month [7] follow-up after the IRM workshop. The process includes raising occupational therapy students' self-efficacy in three areas: the ability to use different therapeutic modes; the ability to identify clients' interpersonal characteristics; and the ability to manage interpersonal events that arise in the treatment. The questionnaire (i.e. Self-Efficacy for Therapeutic Use of Self) that the researchers developed was examined for its reliability and validity with classical test theory [8–10]. However, the instrument has not yet been examined by the analytic approaches based on the modern test theory. Research has suggested that combining modern test theory with the

classical test theory would facilitate a more comprehensive evaluation of an assessment tool [11,12].

Within modern test theory, the Rasch measurement model has been extensively used in assessment development in health- and education-related research [12], as the Rasch techniques provide a mechanism for optimising the test items. Rasch analysis adopted probabilistic test theory, in which the probability of a person to correctly answer one item is a function of the latent variable, and the function represents the intended outcome [13]. Rasch model computation is based on two fundamental assumptions: (1) Within the latent trait, the person with more ability always has a higher possibility of passing any items in the assessment than the person with less ability; (2) The item that is considered harder within the latent trait would be perceived harder by any person than the item that is considered easier. Based on these assumptions, when adopting Rasch analysis in assessment examination, the observed raw data and the expected Rasch responses are compared. If the differences are acceptable (within 40% with the self-report questionnaire), then the data are described to fit the Rasch model [13].

The main benefit of using Rasch analysis is that it transforms ordinal raw data into an interval scale of the latent trait, so that the person's ability and the item's difficulty can be compared linearly. In this perspective, Rasch analysis can be used to evaluate an assessment's construct validity as it generates logits, which are the log-odds probability units, for each item in the questionnaire. Then the latent construct of the questionnaire can be examined [12]. The Equal interval measures generated by Rasch analysis is increasingly being recognised as a possible way to produce data that can inform clinical and educational decision making [14]. The purpose of the current study is to examine the psychometric properties of the Norwegian version of the Self-Efficacy for Therapeutic Use of Self using the Rasch measurement model. The study addresses four specific research questions:

1. How does the rating scale function in the three parts of the questionnaire?
2. Do the items included in each of the three parts demonstrate unidimensionality?
3. Do the participants show expected person response validity in each of the three parts?
4. Do the items in each of the three parts reliably separate the enrolled students into different levels of self-efficacy? Likewise, do the enrolled students

reliably separate the items and confirm the hierarchy levels?

Methods

Study procedures

This study used secondary data analysis with the data retrieved from a larger research project, which included a longitudinal study of occupational therapy students' development of self-efficacy for therapeutic use of self [5-7]. The students self-reported their self-efficacy at four time points (i.e. after the IRM workshop, 3-months, 10-months, and 16-months follow-up). The detailed organisation of the study design and the contents of the IRM workshop is described in a previous article [5].

Participants

A convenience sample of second-year occupational therapy students at two different universities in Norway were invited to participate. The students were informed about the study purposes and procedures; written consent was received from all students before data collection began.

Measurement

Based on the components of Taylor's [4] Intentional Relationship Model (IRM), Yazdani and Tune developed the Self-Efficacy for Therapeutic Use of Self Questionnaire that consists of three parts [5]. Part I is the Self-Efficacy for Therapeutic Mode Use (SETMU) which asks respondents to rate their level of efficacy to apply the six therapeutic modes identified by the IRM. Part II is the the Self-Efficacy for Recognising Interpersonal Characteristics (SERIC) which asks respondents to evaluate their efficacy to identify the 12 interpersonal characteristics presented in clients' communication. Part III is the Self-Efficacy for Managing Interpersonal Events (SEMIE) which asks respondents to rate their level of efficacy for managing the 11 types of interpersonal challenges that may be encountered in clinics. Table 1 lists the items for each part. A 10-point rating scale was used in each part, in which, "1" represents the lowest level (i.e. "I cannot do this at all") and "10" represents the highest level of self-efficacy (i.e. "I am very confident I can do this").

The questionnaire was developed in English and then translated into Norwegian using the forward and backward translation procedures. The forward

Table 1. Psychometric analyses of the Norwegian Self-Efficacy for Therapeutic Use of Self.

Item	Measure	SE	Rasch analysis			
			Infit		Outfit	
			MnSq	Zstd	MnSq	Zstd
Part I: N-SETMU						
Advocate	1.11	0.05	1.22	2.8	1.23	3.0
Instruct	0.49	0.05	1.06	0.9	1.06	0.8
Problem-solving	0.22	0.05	0.84	-2.3	0.86	-2.1
Empathize	-0.43	0.06	1.24	3.1	1.22	2.8
Collaborate	-0.67	0.06	0.97	-0.4	0.98	-0.2
Encourage	-0.71	0.06	0.68	-4.8	0.68	-4.8
Part II: N-SERIC						
Preference for communication style	0.78	0.06	1.18	2.4	1.20	2.6
Response to human diversity	0.31	0.06	0.93	-0.9	0.94	-0.9
Need for control	0.26	0.06	1.03	0.4	1.04	0.5
Orientation towards relating	0.10	0.06	0.83	-2.5	0.82	-2.6
Capacity to assert needs	0.09	0.06	0.95	-0.6	0.95	-0.7
Capacity for reciprocity	-0.01	0.07	0.73	-3.5	0.73	-3.6
Predisposition to giving feedback	-0.08	0.06	0.75	-3.8	0.75	-3.8
Response to change or challenge	-0.14	0.06	0.73	-4.1	0.72	-4.2
Preference for touch ^b	-0.20	0.07	1.54	5.6	1.54	5.4
Predisposition to receiving feedback	-0.24	0.06	1.00	0.0	1.00	0.0
Capacity for trust	-0.26	0.06	1.14	1.8	1.18	2.4
Affect	-0.62	0.06	1.23	3.0	1.24	3.1
Part III: N-SEMIE						
Power dilemmas	0.54	0.05	1.01	0.1	1.00	0.0
Resistance and reluctance	0.28	0.05	0.86	-1.9	0.87	-1.9
Boundary testing	0.27	0.05	0.97	-0.3	0.95	-0.6
Empathic breaks	0.14	0.05	0.93	-0.9	0.93	-1.0
Limitations of therapy	0.11	0.05	0.72	-4.3	0.72	-4.4
Crisis points	0.04	0.05	0.75	-3.8	0.75	-3.8
Expression of strong emotion	-0.09	0.05	1.28	3.7	1.25	3.3
Emotionally charged tasks and situations	-0.13	0.05	0.94	-0.8	0.92	-1.1
Contextual inconsistencies	-0.20	0.05	0.78	-3.3	0.80	-3.0
Non-verbal cues ^b	-0.41	0.05	1.08	1.1	1.08	1.1
Intimate self-disclosures ^b	-0.55	0.05	1.62	7.4	1.61	7.3

Note. ^bInfit statistics > 1.4 associated with Zstd > 2: item misfit.

N-SEMIE: Norwegian version of self-efficacy of managing interpersonal events; N-SERIC: Norwegian version of self-efficacy of recognising interpersonal characteristics; N-SETMU: Norwegian version of self-efficacy of therapeutic mode use.

translation was completed by the last author, who is a licenced occupational therapist, a native Norwegian speaker, fluent in English, and familiar with the terminology of the IRM and self-efficacy. The backward translation was completed by an independent, native English speaker, who had no prior knowledge of the questionnaire when she performed the backward translation. The questionnaire developer compared the backward translation with the original version to ensure that the contents and wordings were clear. No amendments were deemed necessary for the Norwegian version. The questionnaire's construct validity and internal consistency were examined with classical test theory, and the results were satisfactory [8–10].

Data analysis

The Rasch analyses were performed with the Facets, version 3.71.4 [15] and the Winsteps software, version 3.56 [16]. Raw scores from the three parts (N-SETMU, N-SERIC, and N-SEMIE) of the

questionnaire were analysed separately to be consistent with previous studies [8–10]. Rasch analysis performs a logarithmic transformation, which generates an estimation of the participants' level of self-efficacy for using therapeutic mode (i.e. N-SETMU), recognising interpersonal characteristics (i.e. N-SERIC), and managing interpersonal events (i.e. N-SEMIE), respectively, along with the item difficulty calibrations of the same latent traits to be measured.

Due to the nature of repeated measures, in which four time points were included in the data analysis, a random sample dataset was created to estimate the anchor values of the item difficulties from the subjects [17]. The random sample dataset included all the participants across the four time points, but each participant was included only once in the dataset. All time points were equally represented to avoid intra-person dependencies. Then the anchor values of the item difficulty were applied to all the data stacked across four time points, resulting in the total of 400, 399 and 397 observations, in N-SETMU, N-SERIC and N-SEMIE, respectively. By using the random sample dataset to

generate the anchor values of the item difficulty, the intra-person dependencies were no longer a concern because each person was isolated by their multiple completions of the questionnaire *via* the item anchor values [17]. As described in the following section, unidimensionality, rating scale structure, person response validity, item/person separation and the Wright map were evaluated separately in each part of the questionnaire.

Unidimensionality

Unidimensionality was examined by assessing whether the observed responses matched with the expected responses generated from Rasch analysis. In Rasch analysis, every observation contributes to both Infit and Outfit statistics, while the Infit are more sensitive to on-target observations, and Outfit are more sensitive to outliers. Therefore, the goodness-of-fit statistics of the Infit mean-square (MnSq) statistics were used. Any item with an Infit MnSq above 1.4 associated with a standard deviation above 2 was considered misfit [18]. We expected that less than 5% of the items in each part of the measures would fail to meet the criterion [19]. Principal Components Analysis (PCA) of the residuals [20] was used to examine the underlying structure with the criterion that the eigenvalue of the first contrast was less than 3 [21] with the Winsteps software. Additionally, Rasch modelled variances were expected to be close to the empirical variances [21].

Rating scale structure

We examined rating scale functioning by examining whether the rating scale performed consistently across items. The 10-point rating scale within each of the three parts were examined individually based on Linacre's [22] guidelines: (1) At least 10 observations per rating category, (2) the average calibration for each rating category should advance monotonically, and (3) the Outfit MnSq for each rating category should be less than 2. Additionally, the probability curve of each rating category in the three parts were examined separately. Each rating category was expected to have a peak on the curve, which meant that it was the most probable category for some portion of the underlying construct [13].

Person response validity

The person goodness-of-fit statistics derived from Rasch analysis were used to examine person response validity. Participants with an Infit MnSq greater than 1.4 associated with a standard deviation greater than

2.0 were identified as misfit. Generally, we expected that less than 5% of the students would misfit the Rasch model by chance without representing a severe threat to person response validity [21].

Item/person separation

Separation index values range from 0 to infinity; higher values are desired as they indicate better separation ability of the relevant measure [14]. Item separation values were used to examine whether the participants varied enough to confirm the self-efficacy hierarchy in each part of the questionnaire. Linacre [21] suggested that item separation indices of three or greater are desirable. Person separation values were used to examine whether the items in each part were sensitive enough to distinguish students into different self-efficacy levels. Generally, a person separation index of 1.5 is acceptable, 2 is good, and 3 is excellent [23]. The criterion of the separation reliability was set at greater than 0.80 [13].

Wright map

A Wright map (also called person-item map) generated by Rasch analysis provides a visual tool of person-item relationships on an equal interval scale. A Wright map was used to document the hierarchy of items in each part of the questionnaire and to provide clinical guidance related to teaching therapeutic use of self in occupational therapy education.

Results

Unidimensionality

All items except two, (*Preference for touch* in N-SERIC and *Intimate self-disclosures* in N-SEMIE), fit the Rasch model's expectation with acceptable values of Infit MnSq and Zstd (Table 1), which exceeded the pre-determined 5% criterion. Rasch-derived calibration explained 66.2%, 67.5%, and 69.0% of the total variance, which were close to the empirical values of 66.0%, 67.8%, and 69.3% of the total variances explained for the N-SETMU, N-SERIC, and N-SEMIE, respectively. The PCA demonstrated that all the eigenvalue of the first contrast were less than 3: the eigenvalues were 2.2, 1.9, and 2.1 for the N-SETMU, N-SERIC, and N-SEMIE, respectively. The results support the unidimensionality of the three parts of the questionnaire.

Rating scale structure

The lowest two rating categories (i.e. 1 and 2) were the least used categories in all the three parts

Table 2. Category average measure of the Norwegian Self-Efficacy for Therapeutic Use of Self (N-SETUF).

Item	Category average measure ^a									
	1	2	3	4	5	6	7	8	9	10
N-SETMU										
Advocate	-0.05 ^b	0.47 ^b	-0.14 [*]	0.10 [*]	1.02	1.60	2.03	2.64	3.43	5.32
Problem-solving	- ^b	- ^b	-1.22 ^b	-0.43	0.14	0.77	1.64	2.32	2.91	4.47
Instruct	- ^b	-1.94 ^b	-0.29 ^b	-0.20	0.77	0.98	1.72	2.35	3.02	5.41
Encourage	- ^b	- ^b	-1.66 ^b	-0.84 ^b	-0.51	0.28	0.85	1.63	2.45	3.94
Empathize	- ^b	- ^b	-0.64 ^b	-0.32	-0.17	0.90	1.08	1.81	2.39	3.71
Collaborate	- ^b	- ^b	0.10 ^b	-0.46 [*]	-0.51 [*]	0.28	0.85	1.63	2.45	3.94
N-SERIC										
Preference for communication style	- ^b	-2.48 ^b	-1.25	-0.24	0.83	1.82	2.39	3.05	3.68	5.99
Capacity for trust	-3.34 ^b	- ^b	-1.83 ^b	-1.34	0.36	0.98	1.79	2.58	3.43	4.20
Need for control	- ^b	- ^b	-2.40	-0.28	0.50	1.26	2.08	3.03	3.40	5.51
Capacity to assert needs	- ^b	-3.34 ^b	-2.26 ^b	-0.74	0.34	1.33	1.92	2.81	3.49	5.62
Response to change or challenge	- ^b	- ^b	-3.64 ^b	-1.36	0.18	0.97	1.80	2.82	3.53	5.68
Affect	- ^b	- ^b	-2.39 ^b	-0.81	0.02	0.89	1.45	2.27	3.08	4.58
Predisposition to giving feedback	- ^b	- ^b	-2.16 ^b	-1.25	0.07	1.16	1.79	2.81	3.78	5.30
Predisposition to receiving feedback	- ^b	- ^b	-2.59 ^b	-0.42	-0.17	1.08	1.73	2.62	3.36	4.95
Response to human diversity	-3.34 ^b	-2.48 ^b	-1.52 ^b	-0.88	0.50	1.43	2.03	3.04	3.86	5.30
Orientation towards relating	- ^b	-3.34 ^b	-1.16 ^b	-0.91	0.16	1.20	2.15	2.77	3.55	5.40
Preference for touch	- ^b	- ^b	-1.10 ^b	-0.42	0.76	1.43	1.86	2.57	3.01	4.82
Capacity for reciprocity	- ^b	- ^b	-1.82 ^b	-1.22 ^b	0.20	1.20	2.17	2.77	3.84	5.42
N-SEMIE										
Expression of strong emotion	- ^b	-1.38 ^b	-2.16 [*]	-0.98	-0.14	0.42	1.20	1.87	2.47	3.25
Intimate self-disclosures	- ^b	-3.18 ^b	0.88 ^b	-1.05 [*]	-0.59	0.40	0.85	1.52	2.12	2.82
Power dilemmas	- ^b	-3.51 ^b	-1.75	-0.80	0.17	0.85	1.64	2.18	3.18	4.72
Non-verbal cues	- ^b	-0.90 ^b	-2.99 ^{*b}	-1.42 [*]	-0.66	0.26	0.93	1.69	2.43	3.61
Crisis points	- ^b	- ^b	-2.44 ^b	-1.43	-0.36	0.57	1.35	1.88	3.11	3.71
Resistance and reluctance	- ^b	-1.63 ^b	-2.20 [*]	-0.75	-0.22	0.64	1.48	2.10	3.05	4.48
Boundary testing	- ^b	-3.72 ^b	-2.10	-0.81	0.03	0.57	1.35	2.04	2.97	4.38
Empathic breaks	- ^b	- ^b	-1.90	-1.27	-0.19	0.63	1.37	2.06	2.85	3.73
Emotionally charged tasks and situations	- ^b	-2.05 ^b	-2.40 [*]	-1.25	-0.43	0.43	1.23	1.77	2.75	3.70
Limitations of therapy	- ^b	- ^b	-3.23 ^b	-1.42	-0.27	-0.60	1.28	2.21	3.01	4.38
Contextual inconsistencies	- ^b	- ^b	-2.87 ^b	-1.39	-0.63	0.39	1.14	1.88	2.85	3.59

^aThe average measure is expected to increase with category value [22].

^bIndicates its rating category has less than 10 clients.

*Indicates average ability does not ascend with category scores.

(Table 2). It is obvious that the participants had relatively higher self-efficacy than the lowest categories considering that all the items had less than 10 observations in the lowest categories; additionally, none of the participants selected the lowest category (i.e. 1) in N-SEMIE (Table 2). Most rating step thresholds had consistent increased pattern except for two items (i.e. *Advocating* and *Collaborating*) in the N-SETMU, and five items (i.e. *Expression of strong emotion*, *Intimate self-disclosures*, *Non-verbal cues*, *Resistance and reluctance* and *Emotionally charged tasks and situations*) in N-SEMIE (Table 2). Outfit MnSq for each rating category was less than 2 (Table 1). The probability curve of the person-item interaction (Figure 2) showed that the 10-point rating scale did not fully maximise the measurement potentials, as there were too many rating categories. The participants adopted few lower categories, and many middle points had overlapped highest segments.

Person response validity

The person response validity in all three parts exceeded pre-determined criteria (i.e. 5%) with 8.8%,

9.5%, and 7% for the N-SETMU, N-SERIC, and N-SEMIE, respectively.

Item/person separation

The results revealed excellent item and person separation in all three parts. The item separation index was 13.6, 5.9, and 8.1; the person separation index was 2.8, 4.6, and 4.6 for the N-SETMU, N-SERIC, and N-SEMIE, respectively, which all met the desired criteria. The item separation reliability was 0.99, 0.97, and 0.97, and the person separation reliability was 0.89, 0.95, and 0.96 for the N-SETMU, N-SERIC, and N-SEMIE, respectively, which all met the criteria of 0.8.

Wright map

In the Wright map, we examined how well the difficulty of the items matched to the level of self-efficacy of the participants. The results showed that the calibration of the person ability was generally higher than the item difficulty; the mean differences were 1.83 logits for N-SETMU, 1.77 logits for N-SERIC, and 0.9 logits in N-SEMIE (Figure 1). In N-SETMU,

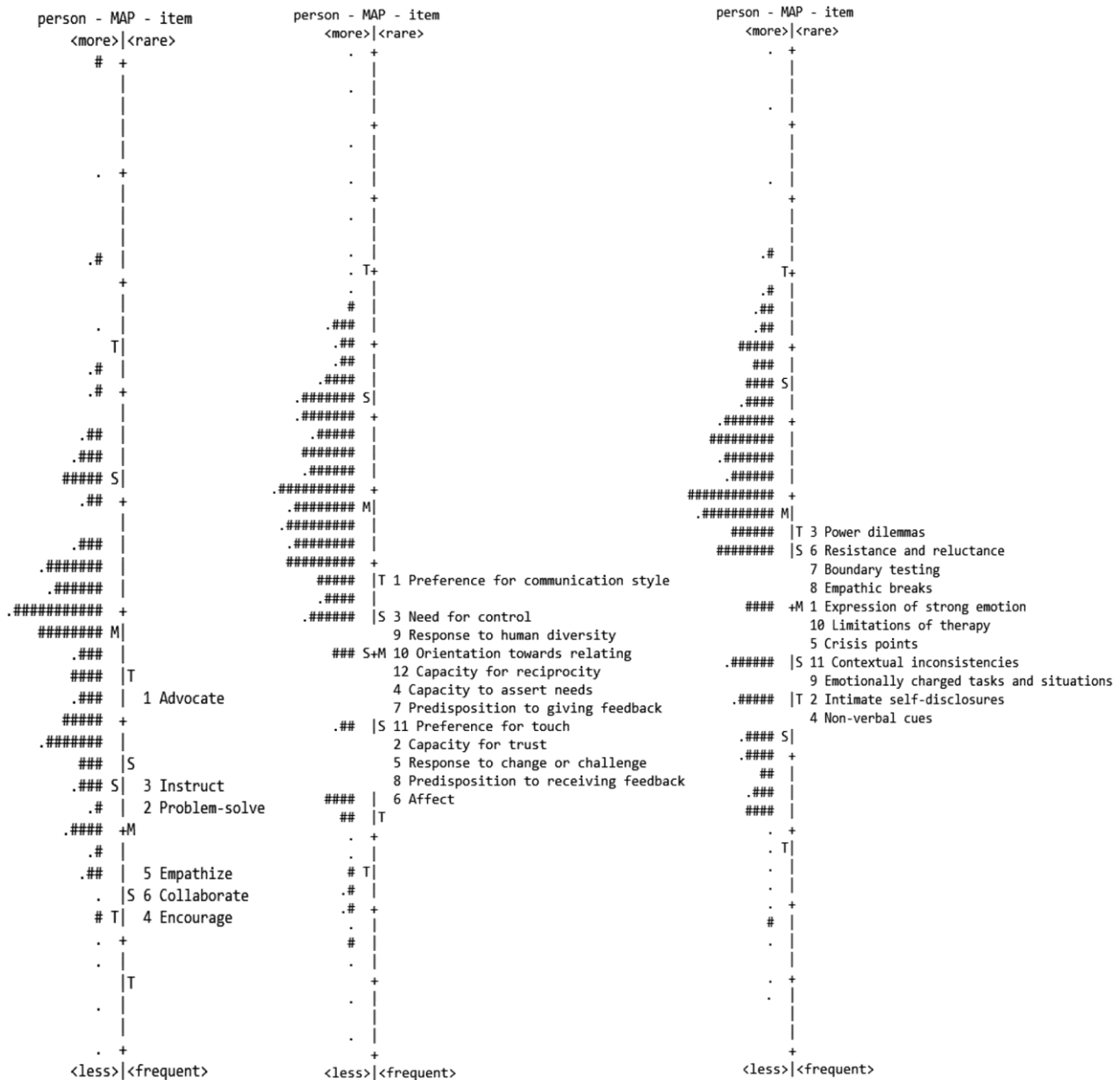


Figure 1. Wright maps of the three parts of the N-SETMU.

participants perceived that the *advocating* mode was the most difficult mode, while the *encouraging* mode was the easiest mode to apply. In N-SERIC, participants had least self-efficacy in recognising clients' *preference for communication style* and had highest self-efficacy in identifying clients' *affect*. In N-SEMIE, participants rated managing *power dilemma* as the hardest interpersonal event, while handling with clients' *non-verbal cues* as the easiest one.

Discussion

The creation of any questionnaire should start with a theory/model about the variables of interest, followed

by rigorous steps to evaluate how well the questionnaire appears to measure the chosen variables based on the theory/model [12]. This study examined the psychometric properties of the newly developed questionnaire measuring self-efficacy for therapeutic use of self, based on the Intentional Relationship Model (IRM) [4] among occupational therapy students, using the Rasch analysis. The Rasch model has solved many of the weaknesses of the classical test theory approach. For example, when using the classical test theory, a completed assessment is required for making results comparisons. Also, the estimates of item difficulty cannot be directly compared unless the estimates come from the same participants or

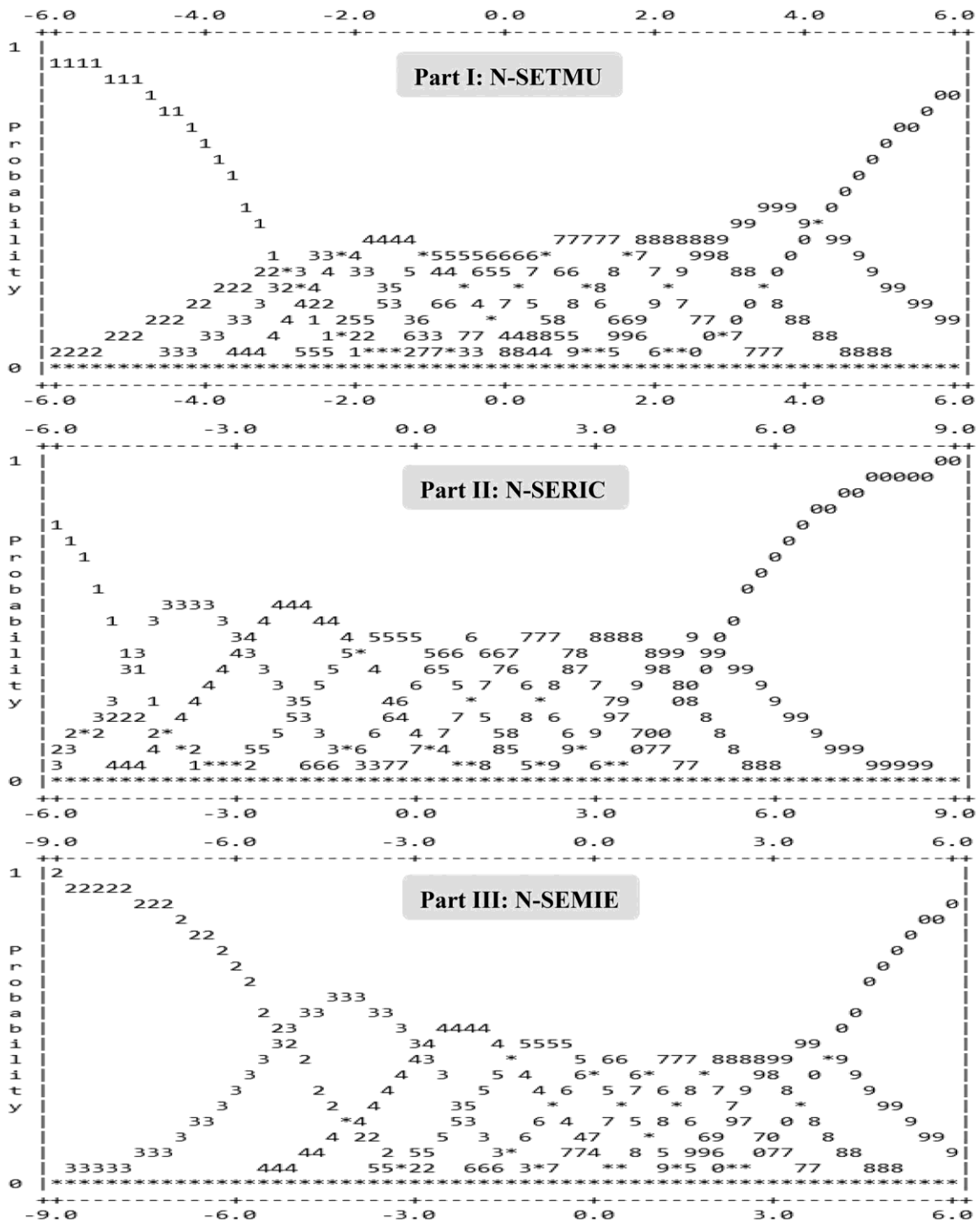


Figure 2. Probability curves of the Norwegian version of Self-Efficacy of Therapeutic Use of Self (part I to part III).

assumptions are made about the comparability of the participants [24]. On the other hand, Rasch analysis yields more comprehensive information, which allows for the connection of observations of participants and items on a linear continuum; therefore, the incomplete assessments with missing data still can be used. Furthermore, information about the structure of the rating scale and the degree to which each item

contributes to the construct is also produced, which complements the deficits of the traditional approaches [24].

Unidimensionality

The results supported the notion that each of the three parts of the questionnaire represented one latent

trait of self-efficacy: therapeutic mode use, interpersonal characteristics recognition, and interpersonal events management, respectively. The results confirmed the construct validity of the questionnaire and were consistent with the previous psychometric studies performed with classical test theory [8–10].

Two items (*Preference for touch* in N-SERIC and *Intimate self-disclosures* in N-SEMIE) did not fit with the Rasch model. With the closer review of the concepts of the IRM, when it comes to clients' *preference for touch*, therapists should be aware of clients' degree of preference for touch and be conscious and knowledgeable about why they might have an aversion to touch. Therapists also should maintain appropriate boundaries and respect clients' preferences for body contact [4]. However, many therapeutic techniques in occupational therapy require therapists to touch clients physically, with various levels of pressure and intensity. Therefore, this item might be rated differently among the study participants.

Regarding the *Intimate self-disclosures*, the IRM emphasises that therapists should respond to those private disclosures in ways that strengthen the therapeutic relationship and build up the feelings of empowerment [4]. However, our student participants may not have enough encounters with clients to feel that they can properly handle this situation with confidence; and therefore, may assume that managing this interpersonal event would involve different constructs or other set of skills. Future studies are needed to examine these two items further.

Rating scale structure

How the self-efficacy measure is divided into rating categories affects the measurement qualities [22]. The current study showed that the lowest rating categories were the least used categories across the three parts, which were consistent with previous findings [8–10]. The 10-point rating scale used in the current questionnaire did not have precisely defined categories, thus obscuring analysis of the rating scales. For example, category 2 was never modal on N-SETMU and N-SERIC (Figure 2), which meant, within these latent traits, at no point that category 2 ever performed as the most likely category to be observed. In other words, the peak of category 2's curve never appeared as a distinct hill but was submerged with nearby rating categories (i.e. category 1 and 3). The current results were consistent with a previous study, in which Fan and Taylor [25] examined the psychometric properties of the Clinical Assessment of Modes

(CAM), an assessment that was developed based on IRM as well. Their results showed that clients could not use the 5-point rating scale in a consistent manner, which resulted in the revision of the questionnaire [26,27]. Linacre [22] suggested combining categories to improve overall quality of the measurement. Concerns about the limited intensity of views of participants [28] also warrant further investigation. Andrew [29] compared the assessments' reliability with scales with a different number of rating categories and concluded that, from 4-point onward, the assessments' reliability increased along with the increased number of categories. Myers [30] also noted that more rating categories would increase the assessments' sensitivity. In the same study, however, Myers [30] suggested that the rating categories should be anchored and labelled appropriately to reduce confusion and remove any ambiguity while more rating options are available. Streiner and Norman [31] stated that in most situations, humans can only discriminate up to seven rating categories. On a related note, Tore and Carstensen [8] considered the scores distribution and suggested using a 7-point scale instead of the 10-point scale. Therefore, future studies may consider adding explicit labels to each rating category or condensing the rating scale. Future studies also should investigate whether participants can interpret rating categories consistently across items.

Person response validity

There was evidence that more than an acceptable proportion of participants (i.e. 5%) were unpredictable in their responses, considering a review of their responses aligns with the difficulty of the items [28]. In other words, the results showed that more than 5% of the participants unexpectedly missed easier items or unexpectedly correctly answered more difficult items. Although fit statistics in the Rasch analysis provide a tool to examine whether participants were idiosyncratic in their responses for various reasons, they did not answer why those participants unexpectedly answered items [32]. One previous study found that most of the work that has examined person misfit is associated with "the dichotomous responses of achievement tests rather than the more complex situation of attitude survey instruments" [33, p.170]. Therefore, participants who misfit to the Rasch model should be investigated further. Although we used fit statistics from Rasch analysis as a quality control for person and item in this study, in the future we anticipate item difficulties will remain and the person

abilities will change when this assessment is applied to other populations. Consequently, we would be “stricter in our application of fit rules to items than to person” [34, p.181]. We suggest that future studies should continue to monitor person response validity across different groups of participants.

Item/person separation

Separation refers to the number of different levels of self-efficacy that the questionnaire can validly identify in the participants (i.e. person separation) or the number of levels of self-efficacy that participants can validly differentiate in the items (i.e. item separation) [21]. The satisfactory person separation indices (2.8–4.6) demonstrated that the items could successfully distinguish subjects into three to five different levels of self-efficacy in the corresponding parts. The excellent item separation indices (5.9–13.6) indicated that the current sample was large enough to confirm precisely the item difficulty hierarchy within the three latent traits.

Wright map

The Wright map is a helpful tool to depict the difficulty of items along with the same linear scale to express the participants’ self-efficacy in terms of different perspectives of the IRM. According to the IRM, there should be no hierarchy of any therapeutic mode; rather, mode use might differ between individuals due to the range of participants’ personal preferences and comfortability. The results showed that the *Encouraging* mode was the least difficult mode to use, which meant that participants had the highest self-efficacy for adopting the *Encouraging* mode. The result was consistent with the exploratory study conducted by Taylor and colleagues [35] among the 563 practicing occupational therapists in the United States. The most difficult mode in the current study was the *Advocating* mode; however, the *Advocating* mode was developed after the original survey [35], making it not comparable. The same study also showed that therapists who worked with clients with more difficult behaviours would tend to use more of the *Instructing* and *Problem-solving* modes, which were the second and the third most difficult modes in our study. In the current study, the participants were second-year occupational therapy students. Although they had participated in clinical observations and level-I fieldworks, they were not assigned to work individually with clients. Instead, licenced

occupational therapists supervised student and client interactions, thus the students might have felt less challenged, possibly resulting in less use of the *problem-solving* and *instructing* modes.

While examining the location of the items in the N-SERIC, the results showed that participants identified recognising clients’ *preference for communication style* as the most difficult item. According to the IRM, identifying communication style includes the understanding of “clients’ preferred amount, nature, and pacing of the interaction that occurs during therapy” [4, p.101]. It might have been very challenging for the study participants to recognise this characteristic due to the complexity involving diverse factors, such as client’s background, neurological impairment, psychiatric symptoms, and other factors [4].

Power dilemmas was categorised as the most difficult item in the N-SEMIE. The current trend in occupational therapy emphasises client-centered therapy [36], which encourages therapists to involve clients in setting up goals as well as desired treatments, so that they have more control in the rehabilitation process. However, occupational therapists undeniably have professional knowledge and resources that may influence clients’ overall functioning; therefore, clients might choose to give up their rights and become dependent on their therapists’ decisions and suggestions. On the other hand, clients might suffer from different levels of discrimination due to their impairments or disabilities. Therefore, they might be much more sensitive than the general population about the power differential with therapists [4]. Hence, it is expected that *power dilemma* is the most difficult item to handle as an interpersonal event for the participants.

In this study, second-year occupational therapy students were invited to fill out the questionnaire. Since they already had completed one year of education in related occupational therapy courses, the students’ education may have resulted in increased self-efficacy for therapeutic use of self. Therefore, very few students selected the lowest score (i.e. category 1) in most of the items. However, social desirability might also skew the data and needs to be considered when interpreting the results. Social desirability is a commonly seen bias in self-report measurements [37–38], in which participants want to present themselves with better performance/outcomes to establish a positive impression and achieve the desired social approval [39–40]. It may be hard for students to identify with the current anchoring statement at the lowest end of the assessment, which states, “I cannot do

this at all.” From a clinical utility perspective, the scales appear to be easier, compared to the students’ current evaluation of their capacity for use of self in therapeutic relationships. However, it is also possible that students tended towards having higher self-efficacy beliefs in the current study due to their lack of experience. These beliefs might change when they encounter relationship difficulties with clients in actual real-life practice.

Limitation and recommendations for future research

The current study had several limitations. First, the selective nature of the convenience sample should be considered when one wants to generalise the study results. Additionally, all enrolled students were second-year occupational therapy students from two universities, and these students may share similar traits in their personality and trainings; therefore, the homogeneity of the sample should be taken into consideration. Additional research is warranted to investigate further the misfit items in N-SERIC and N-SEMIE. Future studies may consider condensing the rating scale or adding explicit anchors to each rating category to examine whether the rating structure improves.

Furthermore, future studies should be conducted to examine the utility of the assessment across various levels of occupational therapy experience. For example, the use of the questionnaire for qualified occupational therapists might require additional investigation as they continue accumulating more years of experience, and their evaluation of self-efficacy for therapeutic use of self might be different than the student evaluations. Last, the current study did not examine differential item function, which could influence the results. Future research should consider examining whether there are systematic differences between gender, experience levels, practice settings, etc.

Conclusion

Rasch analysis provides researchers and educators with analytical tools to conduct detailed analyses in the questionnaire development process. The preliminary outcomes showed that the Norwegian version of the questionnaire, Self-Efficacy for Therapeutic Use of Self, is promising to be used as a reliable and valid tool. This tool could be used as an educational measure when teaching the concepts of therapeutic

relationships, as well as for demonstrating changes in students’ self-efficacy over time while they go through different stages of coursework or in fieldwork practice. However, additional work is required to address the concerns identified in the study, such as the rating scale structure and the misfit items, before it can optimally inform students’ self-efficacy of therapeutic use of self. This questionnaire has sound psychometric properties, and it could be used to tailor occupational therapy educational programmes to meet students’ needs of developing self-efficacy in therapeutic use of self. We expect that the findings from the current study can serve as a guide and reference for occupational therapy educators to measure students’ self-efficacy in therapeutic use of self. The assessment also can help inform the effectiveness of teaching therapeutic use of self and enhance future teaching and learning sessions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- [1] Bandura A. Self-efficacy: the exercise of control. New York (NY): W.H. Freeman and Company; 1997.
- [2] Peloquin SM, Davidson DA. Interpersonal skills for practice: an elective course. *Am J Occup Ther.* 1993; 47(3):260–264.
- [3] Palmadottir G. Client perspectives on occupational therapy in rehabilitation services. *Scand J Occup Ther.* 2003;10(4):157–166.
- [4] Taylor RR. The intentional relationship: occupational therapy and use of self. Philadelphia (PA): F. A. Davis; 2008.
- [5] Hussain RA, Carstensen T, Yazdani F, et al. Short-term changes in occupational therapy students’ self-efficacy for therapeutic use of self. *Br J Occup Ther.* 2018;81(5):276–284.
- [6] Schwank K, Carstensen T, Yazdani F, et al. The course of self-efficacy for therapeutic use of self in Norwegian occupational therapy students: a 10-month follow-up study. *Occup Ther Int.* 2018;2018:2962747.
- [7] Fan CW, Carstensen T, Småstuen MC, et al. Occupational therapy students’ self-efficacy for therapeutic use of self: development and associated factors. *J of Occup Ther Educ.* 2020. Available from: [10.26681/jote.2020.040103](https://doi.org/10.26681/jote.2020.040103)
- [8] Bonsaksen T, Carstensen T. Psychometric properties of the Norwegian self-efficacy for therapeutic mode use (N-SETMU). *Scand J Occup Ther.* 2018;25(6): 475–480.
- [9] Bonsaksen T, Yazdani F, Ellingham B, et al. Psychometric properties of the Norwegian self-

- efficacy for managing interpersonal events (N-SEMIE). *Ergoscience*. 2018;13(3):111-117.
- [10] Ritter VC, Yazdani F, Carstensen T, et al. Psychometric properties of an instrument derived from the Intentional Relationship Model: the self-efficacy for recognizing clients' interpersonal characteristics (N-SERIC). *Open J Occup Ther*. 2018;6(2). Available from: [10.15453/2168-6408.1423](https://doi.org/10.15453/2168-6408.1423)
- [11] Panayides P, Robinson C, Tymms P. The assessment revolution that has passed England by: Rasch measurement. *Br Educ Res J*. 2010;36(4):611-626.
- [12] Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE Life Sci Educ*. 2016;15:1-7.
- [13] Bond T, Fox C. *Applying the Rasch model: fundamental measurement in the human sciences*. 2nd ed. Mahwah (NJ): LEA; 2007.
- [14] Boone WJ, Noltemeyer A. Rasch analysis: a primer for school psychology researchers and practitioners. *Cogent Educ*. 2017;4(1416896):1-13.
- [15] Linacre JM. FACETS Rasch measurement computer program. Version 3.71.4. Chicago (IL): Winsteps.com; 2014.
- [16] Linacre JM. WINSTEPS Rasch measurement computer program. Version 3.56. Chicago (IL): Winsteps.com; 2005.
- [17] Mallinson T. Rasch analysis of repeated measures. *Rasch Meas Trans*. 2011;251(1):1317.
- [18] Wright BD, Linacre JM. Reasonable mean-square fit values. *Rasch Meas Trans*. 1994;8(3):370.
- [19] Patomella A, Caneman G, Kottorp A, et al. Identifying scale and person response validity of a new assessment of driving ability. *Scand J Occup Ther*. 2004;11(2):70-77.
- [20] Hagell P. Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. *OJS*. 2014;04(06):456-465.
- [21] Linacre JM. A user's guide to WINSTEPS[®] MINISTEP Rasch-model computer programs, program manual 4.3.1. Chicago (IL): Winsteps.com; 2018.
- [22] Linacre JM. Optimizing rating scale category effectiveness. *J Appl Meas*. 2002;3(1):85-106.
- [23] Duncan PW, Bode RK, Lai SM, et al. Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale. *Arch Phys Med Rehabil*. 2003;84(7):950-963.
- [24] Wright B. Fundamental measurement for outcome evaluation. *Phys Med Rehabil*. 1997;11(2):261-288.
- [25] Fan CW, Taylor RR. Assessing therapeutic communication during rehabilitation: the Clinical Assessment of Modes. *Am J Occup Ther*. 2016;70(4):7004280010p1-7004280010p10.
- [26] Taylor RR, Fan CW. *Clinical Assessment of Modes—Patient Experiences Version (CAM-E)*. Version 2.0. Chicago (IL): University of Illinois at Chicago, Department of Occupational Therapy; 2015.
- [27] Taylor RR, Fan CW. *Clinical Assessment of Modes—Patient Preferences version (CAM-P)*. Version 2.0. Chicago (IL): University of Illinois at Chicago, Department of Occupational Therapy; 2015.
- [28] Boone WJ, Townsend JS, Staver J. Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: an exemplar utilizing STEBI self-efficacy data. *Sci Ed*. 2011;95(2):258-280.
- [29] Andrew FM, Withey SB. Construct validity and error components of survey measure: a structural modeling approach. *Public Opin Q*. 1984;48:409-442
- [30] Myers JH. *Measuring customer satisfaction: how buttons and other measurement issues*. Chicago (IL): AMA.
- [31] Streiner DL, Norman GR. *Health measurement scales – a practical guide to their development and use*. 4th ed. Oxford: Oxford University Press; 2008.
- [32] Boone WJ, Scantlebury K. The role of Rasch analysis when conducting science education research utilizing multiple-choice tests. *Sci Ed*. 2006;90(2):253-269.
- [33] Karabatsos G. A critique of Rasch residual fit statistics. *J Appl Meas*. 2000;1(2):152-176.
- [34] Boone WJ, Staver JR, Yale MS. Chapter 8. Fit. In: Boone WJ, Staver JR, Yale MS, editor. *Rasch analysis in the human sciences*. The Netherlands: Springer; 2014.
- [35] Taylor RR, Lee SW, Kielhofner G. Practitioners' use of interpersonal modes within the therapeutic relationship: results from a nationwide study. *OTJR*. 2011;31(1):6-14.
- [36] Hinojosa J, Kramer P, Royeen C. *Perspectives in occupational therapy: theories underlying practice*. 2nd ed. Philadelphia (PA): F. A. Davis; 2017.
- [37] Fastame MC, Penna MP. Does social desirability confound the assessment of self-reported measures of well-being and metacognitive efficiency in young and older adults? *Clin Gerontologist*. 2012;35(3):239-256.
- [38] Arnold HJ, Feldman DC. Social desirability response bias in self-report choice situations. *AMJ*. 1981;24(2):377-385.
- [39] Caputo A. Social desirability bias in self-reported well-being measures: evidence from an online survey. *Univ Psychol*. 2017;16(2):245-255.
- [40] Paulhus DL. Two-component models of socially desirable responding. *J Pers Soc Psychol*. 1984;46(3):598-609.