Using Item Response Theory to Evaluate Self-directed Learning Readiness Scale

Rana TH. Momani[1]

[1]Psychology Department, Qassim University, Qassim, Saudi Arabia

Correspondence: Rana TH. Momani, Psychology Department, Qassim University, Qassim, 6666- Buraidah 51452, Saudi Arabia. Tel: 96-562-951-622. E-mail: rana.momani@yahoo.com

Abstract

Item Response Theory becomes one of the most popular methods for instruments development and evaluation methods. This baseline study is a self-directed learning readiness (SDLR) 40 items scale with data from 648 undergraduate psychology female students attending Qassim University in Saudi Arabia through randomized selection to evaluate an SDLR scale at item and scale levels using GRM. Results provide more detailed diagnostic information to modulate the scale. GRM analysis led to the detection of two locally dependent items, one item with low discrimination parameter and 15 model misfit items. The scale often tends to measure low and moderate levels of SDLR. Advanced psychometric evaluations should be made and the SDLR scale must be reviewed based on quantitative and qualitative analysis.

Keywords: graded response model, item response theory, self-directed learning readiness

1. Introduction

Self-directed learning (SDL), which is becoming one of the most important methods in adult education, is defined as the process where learners take the initiative, with or without others assistance, to personalize their learning needs, formulate learning objectives, characterize different resources for learning, and select suitable learning strategies using ongoing evaluation processes for learning objectives (Knowles, 1975).When the concept of self directed learning readiness (SDLR) first appeared, it was defined as the degree to which the learner possesses the attitudes, abilities, and personality characteristics needed for SDL (Wiley, 1983).

Many researchers agree that SDLR's built in assumptions are: adults are ingrained to be self-directed; self-directing could be improved; and, the capability of self-directing could be generalized to many other learning situations (Candy, 1991; Fisher, King, & Tague, 2001; Gugliemino, 1989). Further, researchers also examined the relationship and the effect of SDLR with many educational and psychological variables to find that SDLR could play a main role in students'behavior (Whiley, 1983; O'Kell, 1988; Dyck, 1986). Relatively few scales have been developed to measure students' SDLR within the educational context. Fisher, King, and Tague (2001) developed a scale, recently translated into Arabic by Al Hassoun (2017), which is the exception and might be considered as the most popular scale developed in recent years. Fisher et al. (2001) used their scale in a study of nursing undergraduate students, but did not include any items relating to a specific major, which makes it applicable to any academic field.

Yet, the SDLR scale has not been sufficiently examined psychometrically; few studies have investigated its factor structure. Most researchers who have investigated the psychometric properties of the SDLR scale used Classical Test theory (CTT), thus confining their investigations to the scale's reliability and validity indices to verify the scale's use. Fisher et al. (2001) investigated the psychometric properties of their SDLR scale in a study of 201 nursing undergraduate students. During the scale's development process, they examined the scale's structural validity using principal component analysis and item-total correlation coefficients. The results showed first, that the scale is unidimensional with three main factors: self-management (SM), learning desire (LD), and self-control (SC); and, second, that acceptable values of reliability indices for each scale, which made the scale valid and consistence.

Hendry and Ginns (2009) examined the validity of Fisher et al.'s (2001) SDLR scale in a study of 232 first-year medical students, using exploratory factor analysis (EFA)to define the scale's structure. The results showed that

rather than consisting of three factors, the scale consisted of four: critical self-evaluation, learning self-efficacy, self-determination, and effective organization for learning. Fisher and King (2010) followed with a study that evaluated their scale using confirmatory factor analysis (CFA) on a sample of 227 undergraduate nursing students. Their investigation included using three-one factor congeneric modeling and showed that the best model fit was obtained after the elimination of 11 items – SM1, SM9, SM12, LD7, LD8, LD9, SC1, SC2, SC8, SC10, SC15 – because of their low loadings.

Torabi et al. (2013) examined Fisher et al.'s (2001) SDLR scale among preliminarily schoolteachers in Esfahan. Results showed that the scale approximately fit using CFA, and they concluded that it could be used to evaluate SDLR among teachers. Williams and Brown (2013) investigated Fisher et al.'s (2001) SDLR scale construction using several CFA models on a sample of 233 undergraduate Australian students. The results showed that the model of 40 items with three factors was not a fit for the data, while the 36 items with the four factors model fit better than the 29 items with the three factors model.

As the concept of SDL readiness currently receives more attention, the above-mentioned research only attempted to use CTT and factor analyses methods to verify the SDLR scale. The scientific revolution and the existence of several research instruments means the evaluation of research instruments is an essential need especially for those that are frequently used, such as Fisher et al.'s SDLR scale. Although this scale has yet to be investigated using advanced psychometric analysis methods, the current study aims to evaluate the SDLR scale using IRT, one of the most important approaches to evaluate and develop scales because of its accuracy in item and personal levels.

2. Research Method

2.1 Participants

The present study involved the participation of 636 female undergraduate psychology students, with an average age of 19.91(sd= 0.99), in the faculty of education at Qassim University in Saudi Arabia during the academic year 2016/2017: participation was on a voluntary basis.

2.2 Self Directed Learning Readiness (SDLR)

The participants of the present study were subjected to the 40 items SDLR scale developed by Fisher et al. (2001), translated into Arabic by AlHassoun (2017). The SDLR contained three parts designed to estimate three dimensions of the SDLR: self-management, learning desire, and self-control. Each part consists of 13, 12, and 15 items.

2.3 Research Procedure

The SDLR scale was administered to the study participants, each participant required approximately 25 minutes to complete the scale.

2.4 Statistical Procedure

The SDLR scale has a graded polytomous response format and Samejima's graded response model (GRM) is considered the optimal IRT model to use. The GRM was designed for ordinal polytomous items, and is considered a generalization for the 2–Parameter model, which expresses the behavior of an item i by its discrimination parameter (a_i), and a set of threshold parameters ($b_{ij} = 1, …… ., m$) located between the continuous categories of a polyatomic items ($j = 1, …… ., m$) (Attorresi, Abal, Galibert, Lozzia, and Aguerry, 2011; Samejima, 1969).

The GRM, like all IRT models derives the probability of the responses of each item as a function of the latent trait θ and item parameters. It is also an estimate of the cumulative probability of responding in a category or above. This probability can be plotted for each category that produces categorical response curves (CRC) (DeMaes, 2010).

Baker (2001) pointed that GRM, like most IRT models, relies on two main assumptions:

1) Unidimensionality – all items belong to a common construct that influence the item responses.

2) Local independence – all elements of the item vector for respondents are independent from each other.

3. Research Findings

3.1 Basic Statistics

Total scores for the SDLR and its dimensions were computed. Then, the means and standard deviations were extracted, as shown in Table 1. The results illustrated that the sample performance was best for self-control. This result can be attributed to how Muslims raise their children, with commitment, avoidance of rage, respect toward others, kept away from taboos, and other ethics ordered by the Islamic religion. In other words, this result was caused by the effect of Islamic rules that control Muslims' internal and external behavior, which leads to Muslims' high self-control.

Table 1. Basic statistics for the scale dimensions

	Self-Management	Learning Desire	Self-Control	Total
Mean	47.14	46.03	59.47	152.64
SD	8.19	6.31	8.53	19.46

The means and standard deviations of each item were also calculated, as shown in Table 2, for further clarification of the participants' performances based on the SDLR scale.

Table 2. Statistics for SDLR scale items

SM1	SM2	SM3	SM4	SM5	SM6	SM7	SM8	SM9	SM10	SM11	SM12	SM13		
3.38	3.77	3.75	2.72	3.54	3.46	3.58	3.44	3.50	3.87	4.11	3.96	4.24		
0.95	1.10	1.17	1.19	1.13	1.23	1.19	1.42	1.12	1.14	1.07	1.21	1.01		
LD1	LD2	LD3	LD4	LD5	LD6	LD7	LD8	LD9	LD10	LD11	LD12			
4.19	4.14	3.80	4.27	2.86	2.97	3.96	3.92	4.02	4.25	4.27	3.53			
1.07	1.07	1.28	1.11	1.27	1.22	1.16	1.13	1.09	1.02	1.06	1.27			
SC1	SC2	SC3	SC4	SC5	SC6	SC7	SC8	SC9	SC10	SC11	SC12	SC13	SC14	SC15
4.20	4.11	4.54	4.02	3.60	3.89	3.71	3.68	3.96	3.76	3.85	4.09	4.16	4.19	3.69
1.03	1.04	0.83	1.02	1.17	1.19	1.20	1.14	1.20	1.14	1.11	1.01	0.96	1.00	1.14

Note. SMi: Self-management items; LDi: Learning desire items; SCi: Self Control items.

As shown in Table 2, the item labeled SC3 (I am responsible for my own decisions) received the highest mean value. But, responses were less varied across sample respondents, which could be explained by the respondents' age: respondents were adolescents and tend toward independent decision making.

3.2 Reliability

The values of the alpha coefficient were computed for each dimension score and the overall scale, as shown in Table 3. The results illustrated that the SDLR scale has a moderate value of reliability indices. The learning desire item had the lowest alpha value, which could be explained by the number of items for this dimension.

Table 3. Alpha coefficient values for SDLR scale and its dimensions

	Self-Management	Learning Desire	Self-Control	Total
Alpha	0.83	0.71	0.83	0.89
N	13	12	15	40

3.3 Correlations

The correlations between scale dimensions and total score were computed, as shown in Table 4.

Table 4. The correlations between SDLR scale dimensions and total score

	Self-Management	Learning Desire	Self-Control	Total
Self-Management	1.00	0.58**	0.67**	0.88**
Learning Desire		1	0.62**	0.82**
Self-Control			1	0.90**
Total				1

The correlations showed significant relationships among the three dimensions of SDLR, which confirmed that each dimension represents a component that correlates with the others. The correlations also showed moderate and significant relationships (p<0.01) between SDLR dimensions, and a highly significant relationship with the total score of the SDLR, which proved its consistency.

3.4 Checking Model Assumptions

Unidimensionality, which is a primary assumption for GRM, was checked using both EFA and CFA. EFA yielded a general factor with 22.85% initial variance, and the ratio of initial variance for the first and second factors was 4.28. As Reckase (1979) suggested, if the first factor has a variance that exceeds 20%, and if the ratio of initial variance for the first and second factors exceeds 2, these results confirm the scale's unidimensionality.

CFA was also conducted for the 40 items of SDLR using a one general factor model with three subscales model. The goodness of fit (GOF) indices were computed: goodness of fit index (GFI) =0.90, comparative fit index (CFI) =0.90, incremental fit index (IFI) =0.90, and root mean square error approximation (RMSEA) =0.037. While all GOF indices had acceptable values, the model of one factor model was a good fit with the data, which is the second piece of evidence for unidimensionality.

Local independence was examined by checking the discrimination parameter a_i. Items with high slope parameter (e.g., > 4)refer to potential violations. The values of local independence chi square (LDχ^2) were also used for an additional assessment for local independence. An item with LD$\chi^2 = 10$ suggests serious violations for item independency (Nguyen, Han, Kim, & Chan, 2015). The results showed that all slope parameters had values less than 4, and the values of LDχ^2 for items pairs showed significant violation in items independency for two items: LD1 (I want to learn new information) and LD2 (I enjoy learning new information), which seem to have nearly the same meaning.

3.5 IRT Analyses of SDLRScale

The analysis of the SDLR items using GRM, which is one of the IRT models for polytomous items, were computed using the Bock-Aitikin method implemented in IRTPRO4 software (Cai, Thissen, & du Toit, 2011).

Item parameters for the SDLR scale were computed using GRM, as shown in Table 5, whereb_i's represents the ability level with an associated probability of 0.5 to respond to the above threshold. The location parameter$a_i$represent the discrimination power of each item between respondents with high and low ability scores.

Table 5. Item parameters for SDLR scale

Label	a	s.e.	b1	s.e.	b2	s.e.	b3	s.e.	b4	s.e.
SM1	1.19	0.11	-3.50	0.33	-1.94	0.17	0.34	0.10	1.54	0.17
SM2	1.20	0.11	-3.25	0.30	-2.01	0.18	-0.43	0.08	0.59	0.11
SM3	1.22	0.11	-2.62	0.23	-1.61	0.14	-0.21	0.08	0.88	0.12
SM4	1.21	0.11	-1.64	0.14	-0.20	0.08	0.92	0.12	1.95	0.19
SM5	1.34	0.12	-2.74	0.23	-1.45	0.12	-0.17	0.07	0.92	0.12
SM6	1.34	0.12	-2.28	0.19	-1.27	0.11	-0.03	0.08	0.81	0.11
SM7	1.34	0.12	-2.51	0.21	-1.50	0.13	-0.17	0.07	0.72	0.10
SM8	1.01	0.11	-2.24	0.23	-1.11	0.13	-0.10	0.09	0.62	0.12
SM9	1.31	0.12	-3.03	0.27	-1.37	0.12	-0.04	0.08	0.96	0.12
SM10	1.23	0.12	-3.07	0.29	-1.91	0.17	-0.65	0.09	0.33	0.09
SM11	1.34	0.13	-3.15	0.29	-2.19	0.19	-0.99	0.10	-0.04	0.08
SM12	1.19	0.12	-2.83	0.27	-1.95	0.18	-0.74	0.09	0.01	0.08
SM13	1.22	0.12	-3.91	0.40	-2.51	0.23	-1.24	0.12	-0.28	0.08
LD1	0.97	0.11	-4.13	0.47	-3.00	0.32	-1.17	0.14	-0.39	0.10
LD2	0.94	0.11	-4.49	0.52	-2.96	0.32	-1.08	0.14	-0.25	0.10
LD3	0.81	0.10	-3.39	0.40	-2.25	0.26	-0.81	0.13	0.31	0.12
LD4	0.58	0.10	-5.76	0.96	-4.19	0.69	-2.43	0.40	-0.97	0.20
LD5	1.08	0.10	-1.69	0.16	-0.60	0.09	0.76	0.12	1.83	0.19
LD6	0.83	0.09	-2.06	0.23	-0.68	0.12	1.04	0.16	2.57	0.30
LD7	1.26	0.12	-2.95	0.27	-1.90	0.17	-0.78	0.09	0.09	0.08
LD8	1.11	0.11	-3.24	0.32	-2.21	0.21	-0.83	0.10	0.32	0.10
LD9	1.19	0.12	-3.42	0.34	-2.24	0.20	-0.85	0.10	0.12	0.09
LD10	0.93	0.11	-4.64	0.55	-3.21	0.36	-1.47	0.17	-0.44	0.10
LD11	0.49	0.09	-7.52	1.43	-5.26	0.98	-2.77	0.52	-0.94	0.23
LD12	0.50	0.08	-5.14	0.84	-2.65	0.44	-0.40	0.17	1.52	0.30
SC1	1.69	0.16	-3.28	0.30	-1.98	0.15	-0.88	0.08	-0.24	0.06
SC2	0.97	0.11	-4.83	0.57	-2.84	0.30	-1.11	0.13	-0.05	0.09
SC3	0.89	0.12	-5.65	0.78	-4.16	0.53	-2.27	0.28	-1.21	0.16
SC4	1.02	0.11	-4.79	0.55	-2.70	0.27	-0.92	0.11	0.24	0.10
SC5	1.34	0.13	-2.53	0.22	-1.53	0.13	-0.27	0.07	0.78	0.11
SC6	1.69	0.15	-2.32	0.18	-1.51	0.11	-0.54	0.06	0.14	0.07
SC7	1.25	0.12	-2.73	0.25	-1.65	0.15	-0.39	0.08	0.53	0.10
SC8	1.29	0.12	-2.87	0.26	-1.71	0.15	-0.38	0.08	0.69	0.11
SC9	1.08	0.12	-3.03	0.31	-2.15	0.21	-0.85	0.10	0.05	0.09
SC10	1.35	0.13	-2.90	0.26	-1.68	0.14	-0.41	0.07	0.52	0.10
SC11	1.20	0.12	-3.56	0.35	-1.91	0.17	-0.63	0.08	0.43	0.10
SC12	1.33	0.13	-3.84	0.39	-2.31	0.20	-0.94	0.09	0.07	0.08
SC13	0.90	0.10	-5.20	0.64	-3.66	0.41	-1.39	0.16	0.02	0.10
SC14	1.46	0.14	-3.80	0.39	-2.19	0.18	-1.01	0.09	-0.14	0.07
SC15	1.68	0.15	-2.35	0.18	-1.47	0.11	-0.32	0.06	0.57	0.09

Note. b_i = Threshold Parameter.

3.6 Threshold Parameter

The polytomous nature of SDLR scale items with five possible responses means that each item has four response threshold parameters: b_1, b_2, b_3, and b_4. According to GRM, these thresholds depict the trait level with a 50% chance of scoring at or above a scale response, and they express important information about items (Steinberg &Thissen, 1995).

While the SDLR scale is self-reported, and measures how respondents feel about themselves and self-directed learning readiness, it is also desirable that scale items introduce good information on a wide range of the latent trait.

Table 5 shows the threshold parameters for the SDLR scale wherein each item in the scale covered a certain range in the latent continuum: some items had relatively low threshold parameters (e.g., LD4: I enjoy a challenge, LD10: I learn from my mistakes), while others had relatively high threshold parameters (e.g., LD5: I enjoy studying, SM4: I set strict time frames). Nevertheless, the results in Table 5 showed that items thresholds were consistent, and each item covered an acceptable range of the latent trait, but mostly in the moderate and low trait level with some expectations for items SM1, SM4, LD5, LD6, and LD12. Item SC3 (I am responsible for my own decisions/ actions) covered a low limited ability level of the latent trait, however, which was below -1 for all categories, which could be explained by its generality (not specified in self directer learning context). The rest of the SDLR scale tended to be at the low-mid level of the latent trait continuum as most of the $b's$ had negative values, which is less than the midpoint of the latent trait continuum.

3.7 Discrimination (Slope) Parameters

Results of Table 5 show that item slopes a_i range from 0.49 to 1.69. Items LD4 (I enjoy a challenge), LD11 (I need to know why), and LD12 (When presented with a problem I can't resolve, I will ask for assistance), had low discrimination parameters, which could be explained by items content, which is a type of generalization not necessarily related to self directed learning ability. Items SC1 (I prefer to set my own goals), SC6 (I prefer to set my own learning goals), SC10 (I have high personal expectations), SC14 (I have high belief in my abilities), and SC15 (I prefer to set my own criteria on which to evaluate my performance) had high discrimination parameters, which could be explained by their content that is well specified and related to self directed learning readiness. The remaining items had acceptable moderate discrimination parameters, according to Baker (2001).

Category response functions (CRF) were also extracted, as shown in Figure 1. CRF represents the variety of score probabilities as a function of the latent trait, they describe the probability that the respondent will respond to an item with 1, 2, 3, 4, or 5.
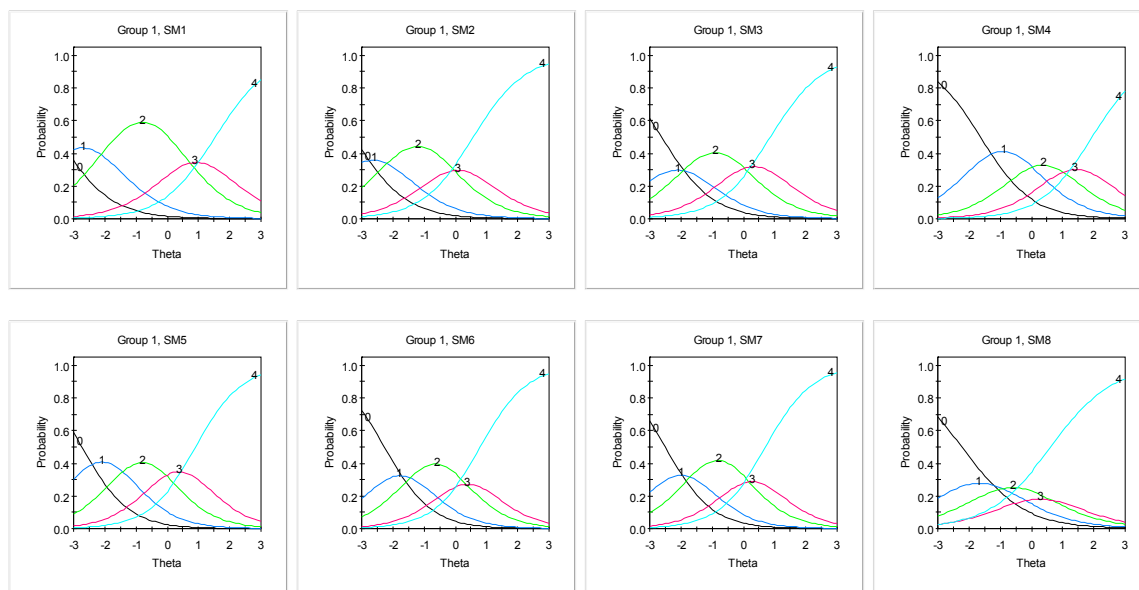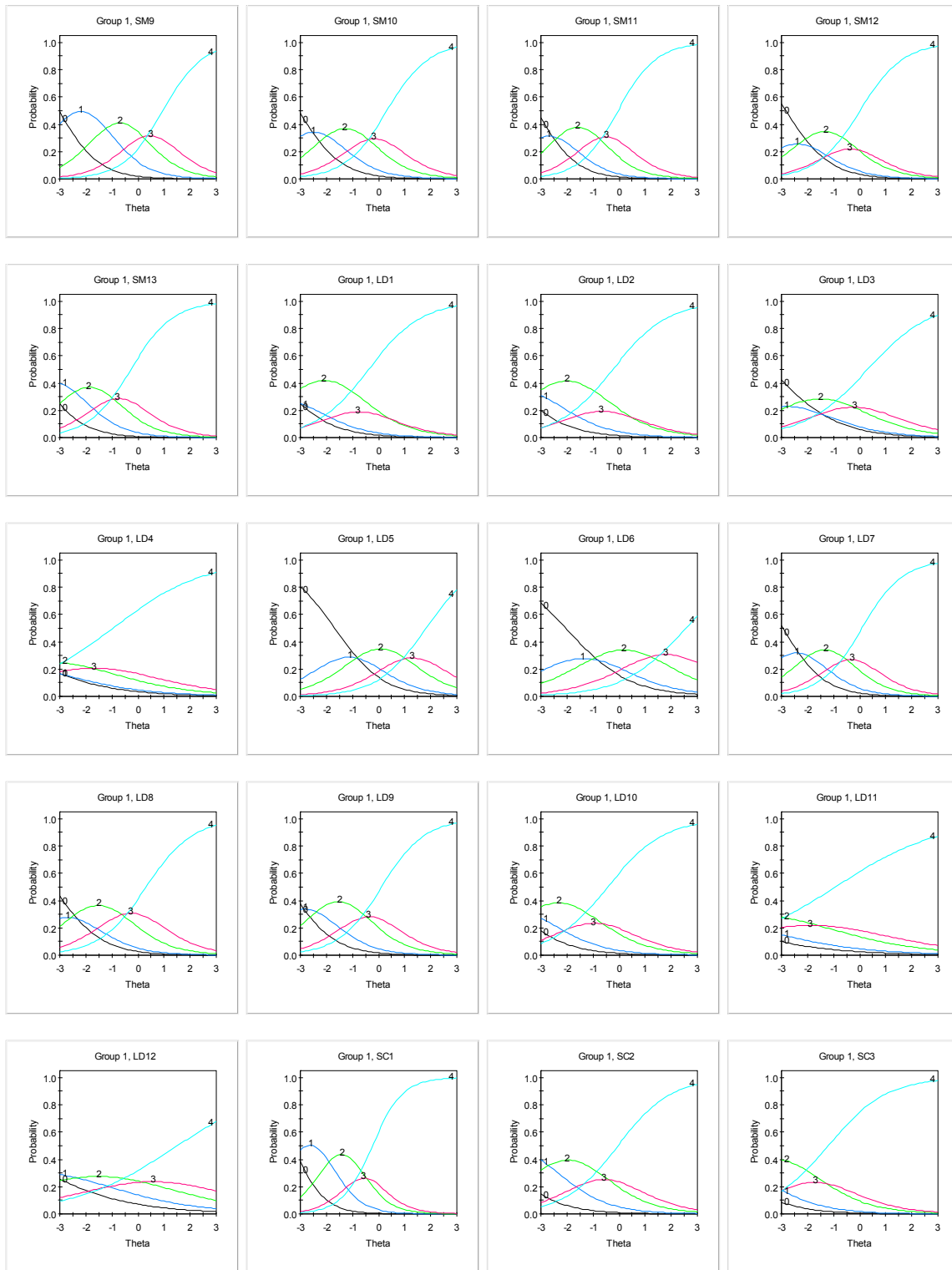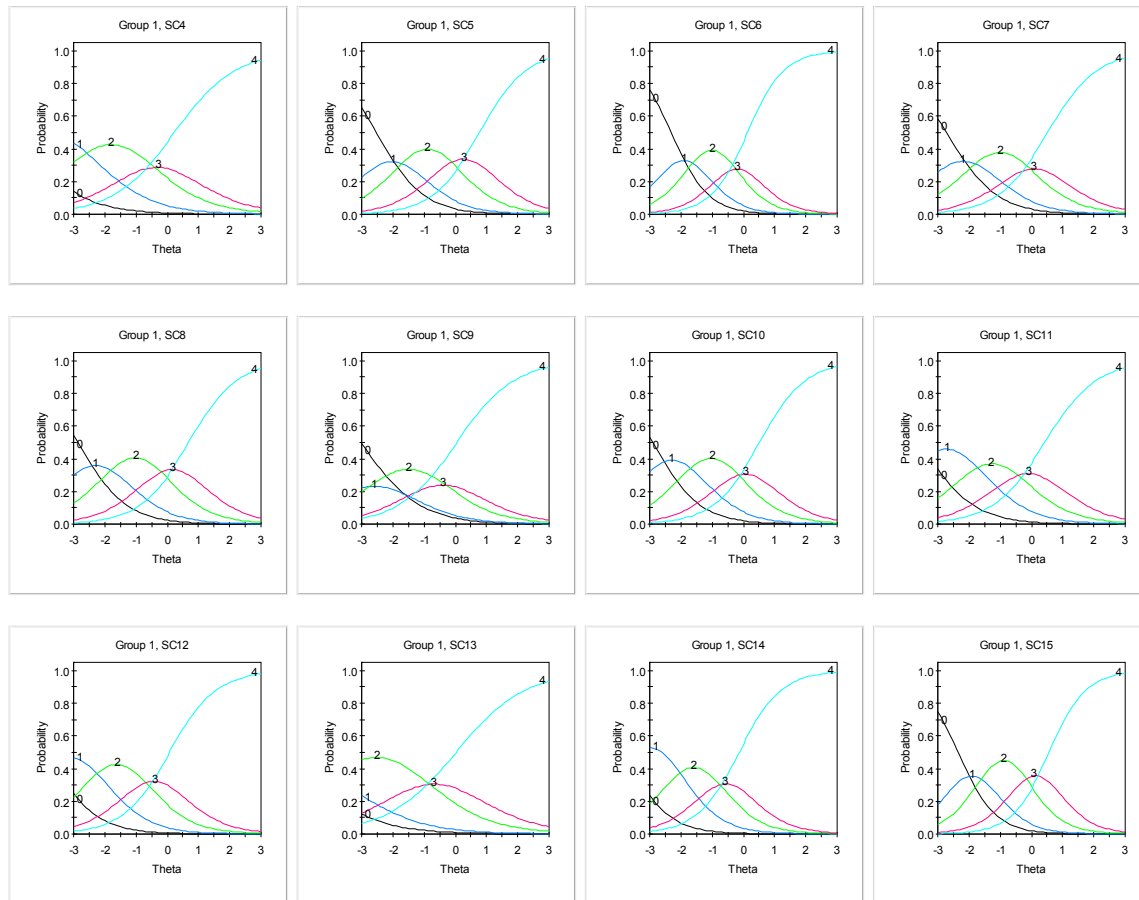
Figure 1. CRFs for SDLR scale

As shown in Figure 1, some items had low peaks and high overlap between category functions, such as item LD4, which led to low discrimination parameters. While others had high peaks and less category functions overlap, such as item SC6, which led to high discrimination parameters.

The information introduced by each item and the whole scale were computed, as shown in Table 6.

Table 6. SDLR scale items information

Label/θ	-2.80	-2.40	-2.00	-1.60	-1.20	-0.80	-0.40	0.00	0.40	0.80	1.20	1.60	2.00	2.40	2.80
SM1	0.41	0.41	0.41	0.39	0.38	0.37	0.38	0.40	0.42	0.42	0.42	0.39	0.34	0.28	0.21
SM2	0.43	0.44	0.44	0.44	0.43	0.44	0.44	0.44	0.42	0.39	0.33	0.26	0.19	0.13	0.09
SM3	0.40	0.44	0.45	0.46	0.45	0.45	0.46	0.46	0.45	0.42	0.38	0.32	0.24	0.17	0.12
SM4	0.24	0.31	0.37	0.41	0.43	0.44	0.45	0.46	0.46	0.46	0.46	0.44	0.41	0.36	0.29
SM5	0.49	0.52	0.54	0.54	0.54	0.54	0.55	0.55	0.54	0.52	0.46	0.37	0.28	0.19	0.12
SM6	0.41	0.49	0.53	0.55	0.55	0.55	0.55	0.55	0.54	0.51	0.44	0.35	0.25	0.17	0.11
SM7	0.46	0.51	0.54	0.55	0.54	0.54	0.55	0.55	0.53	0.49	0.42	0.33	0.23	0.15	0.10
SM8	0.25	0.28	0.30	0.32	0.32	0.33	0.32	0.32	0.31	0.28	0.25	0.21	0.17	0.13	0.09
SM9	0.48	0.48	0.49	0.50	0.51	0.51	0.52	0.52	0.52	0.50	0.45	0.37	0.28	0.20	0.13
SM10	0.45	0.46	0.47	0.47	0.47	0.47	0.47	0.46	0.42	0.37	0.29	0.22	0.15	0.10	0.07
SM11	0.54	0.55	0.56	0.55	0.56	0.55	0.54	0.50	0.43	0.34	0.24	0.16	0.10	0.06	0.04
SM12	0.41	0.43	0.44	0.44	0.44	0.44	0.43	0.40	0.36	0.29	0.22	0.16	0.11	0.07	0.05

SM13	0.45	0.46	0.46	0.46	0.46	0.46	0.43	0.39	0.33	0.25	0.18	0.12	0.08	0.05	0.03
LD1	0.29	0.29	0.29	0.29	0.29	0.29	0.27	0.25	0.21	0.18	0.14	0.11	0.08	0.06	0.04
LD2	0.27	0.27	0.27	0.27	0.27	0.27	0.26	0.24	0.21	0.18	0.15	0.11	0.09	0.06	0.05
LD3	0.20	0.21	0.21	0.21	0.21	0.21	0.20	0.20	0.19	0.17	0.15	0.13	0.11	0.09	0.07
LD4	0.11	0.10	0.10	0.10	0.10	0.09	0.09	0.08	0.07	0.07	0.06	0.05	0.04	0.04	0.03
LD5	0.21	0.26	0.30	0.34	0.35	0.36	0.36	0.36	0.36	0.36	0.36	0.34	0.32	0.28	0.23
LD6	0.16	0.18	0.20	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.21	0.20	0.19
LD7	0.46	0.49	0.50	0.50	0.50	0.50	0.49	0.46	0.41	0.34	0.26	0.18	0.12	0.08	0.05
LD8	0.37	0.38	0.38	0.38	0.38	0.38	0.38	0.37	0.34	0.30	0.25	0.19	0.14	0.10	0.07
LD9	0.43	0.44	0.44	0.44	0.44	0.44	0.43	0.41	0.37	0.31	0.24	0.18	0.12	0.08	0.05
LD10	0.27	0.27	0.27	0.27	0.26	0.26	0.24	0.22	0.19	0.16	0.13	0.10	0.07	0.05	0.04
LD11	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.03	0.03
LD12	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.07	0.07	0.07	0.06	0.06
SC1	0.81	0.82	0.84	0.85	0.87	0.88	0.85	0.74	0.55	0.36	0.21	0.12	0.06	0.03	0.02
SC2	0.28	0.28	0.29	0.29	0.29	0.29	0.28	0.26	0.24	0.21	0.17	0.13	0.10	0.07	0.05
SC3	0.24	0.24	0.24	0.24	0.23	0.21	0.18	0.15	0.12	0.10	0.07	0.06	0.04	0.03	0.02
SC4	0.31	0.31	0.31	0.31	0.32	0.32	0.32	0.31	0.29	0.25	0.21	0.17	0.13	0.09	0.07
SC5	0.46	0.52	0.54	0.55	0.55	0.55	0.55	0.55	0.54	0.50	0.43	0.34	0.25	0.17	0.11
SC6	0.63	0.78	0.86	0.88	0.88	0.88	0.88	0.84	0.72	0.54	0.35	0.20	0.11	0.06	0.03
SC7	0.44	0.47	0.48	0.49	0.49	0.49	0.49	0.48	0.46	0.41	0.34	0.26	0.19	0.13	0.08
SC8	0.47	0.50	0.51	0.51	0.51	0.51	0.51	0.51	0.49	0.46	0.39	0.31	0.22	0.15	0.10
SC9	0.35	0.36	0.36	0.37	0.37	0.36	0.35	0.34	0.30	0.26	0.21	0.16	0.11	0.08	0.05
SC10	0.52	0.54	0.55	0.55	0.55	0.56	0.56	0.55	0.53	0.47	0.38	0.28	0.19	0.12	0.08
SC11	0.42	0.43	0.44	0.44	0.45	0.45	0.45	0.44	0.41	0.37	0.30	0.23	0.17	0.11	0.08
SC12	0.51	0.52	0.53	0.53	0.53	0.54	0.53	0.50	0.44	0.36	0.27	0.18	0.12	0.07	0.04
SC13	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.23	0.21	0.19	0.16	0.13	0.10	0.08	0.06
SC14	0.60	0.62	0.64	0.65	0.65	0.66	0.64	0.58	0.47	0.35	0.23	0.14	0.09	0.05	0.03
SC15	0.63	0.78	0.85	0.86	0.85	0.84	0.85	0.86	0.82	0.72	0.55	0.36	0.22	0.12	0.06
test	16.25	17.25	17.78	17.99	18.03	18.02	17.85	17.27	16.05	14.18	11.88	9.49	7.34	5.57	4.22
s.e.	0.25	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.25	0.27	0.29	0.32	0.37	0.42	0.49

As well as item and scale, information functions were plotted for the SDLR scale for each trait level, as shown in Figure 2.
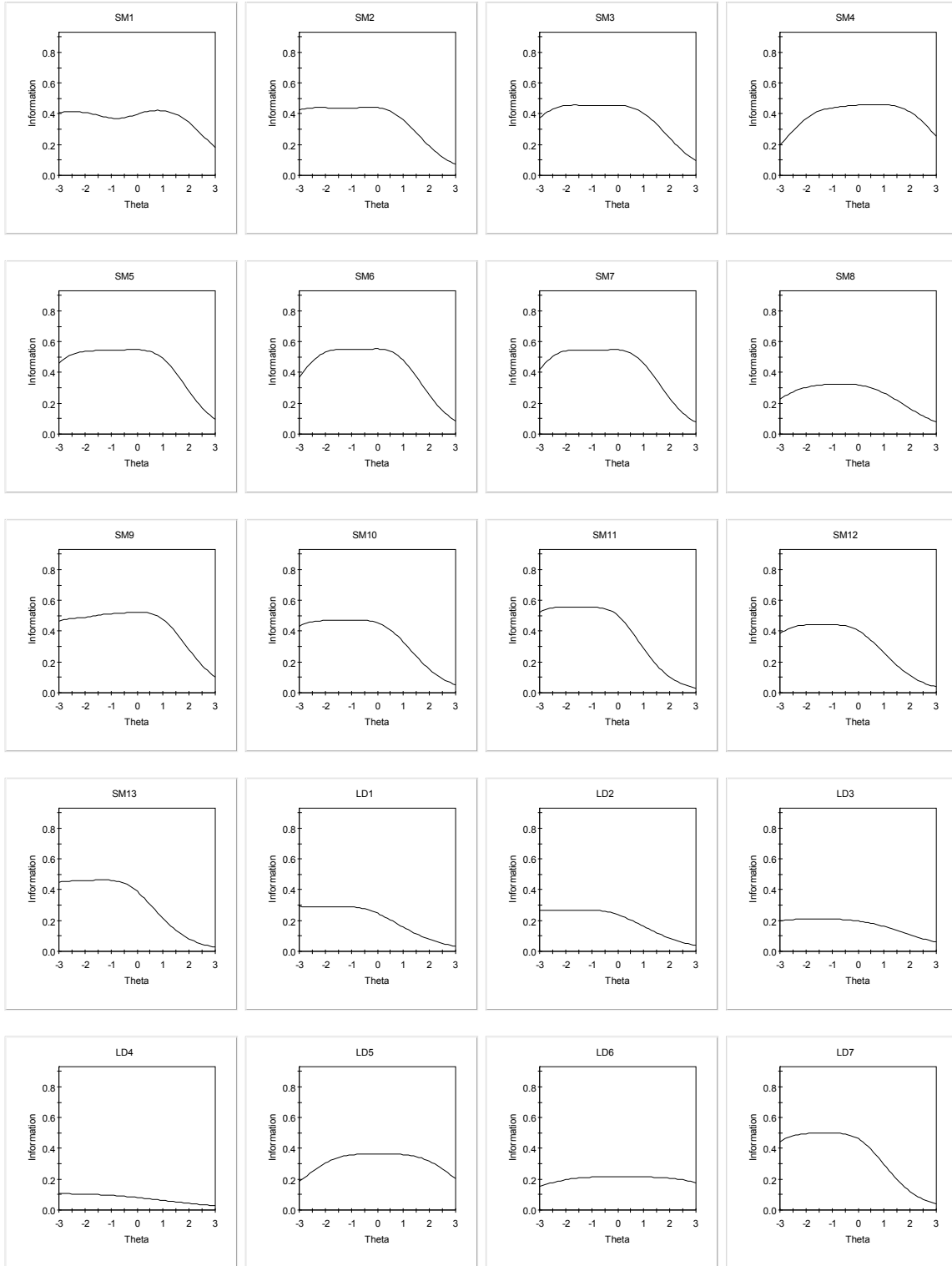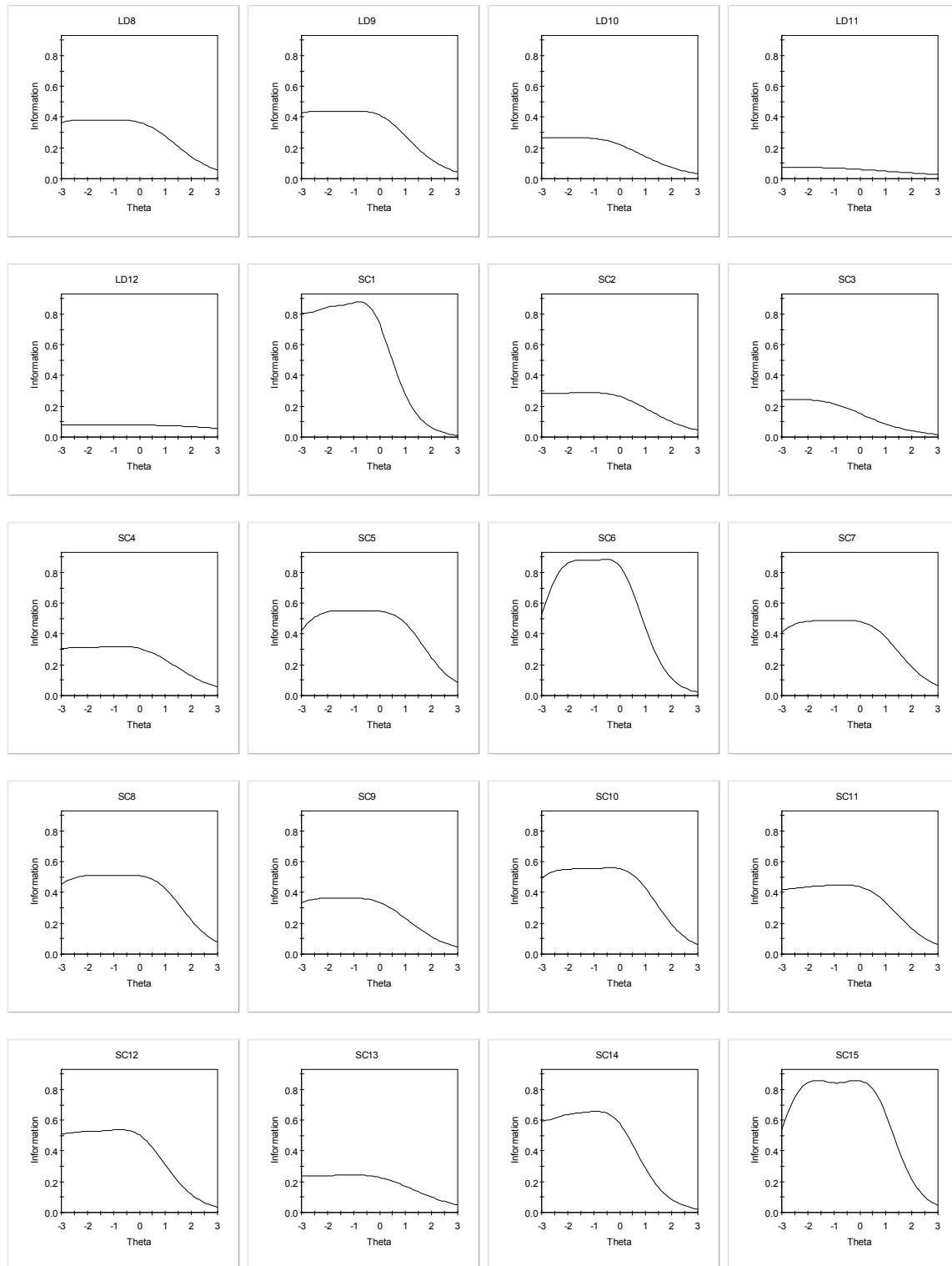
Figure 2. Item Information Functions for SDRL scale

As shown in Figure 2, most of the SDLR scale items had relatively high information at low ability levels of the trait, exclusively some items introduced good information along the ability continuum, such as items SM4 and LD5, while other items were quite weak and introduced low information along the ability continuum, such as items LD11 and LD12. This result was associated with low discrimination power for these two items, which made them unable to provide good information across the ability continuum.

3.8 Test Information Function

The test information function (TIF) is simply the sum of the item information at each ability level (Baker, 2001). It is considered a useful feature to estimate the quality of the test as a measure of a particular latent trait. Figure 3 shows the test information function for the SDLR scale. As shown in Figure 3, the total information for the SDLR scale has good reliability for low and moderate ability levels of the trait. This reliability becomes lower for ability levels above 1.2, and the marginal reliability index was 0.93, which indicates that the SDLR scale is reliable.
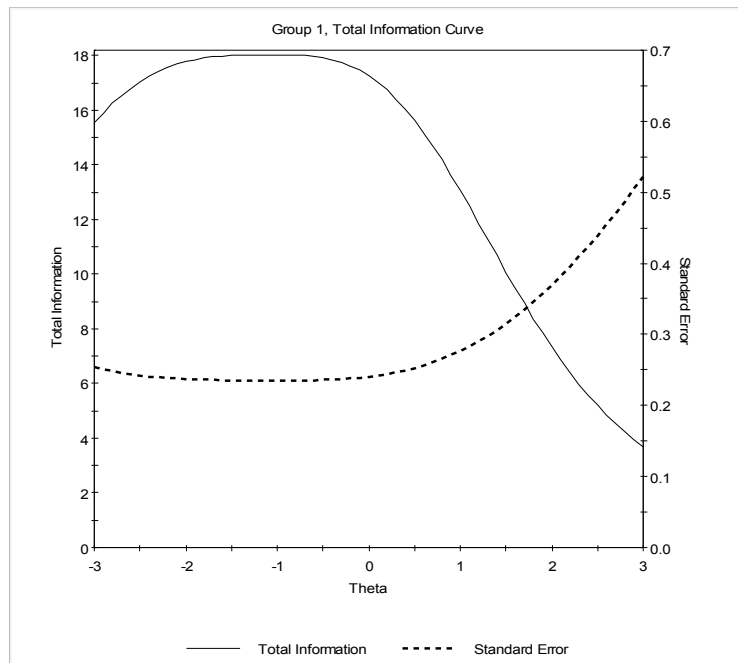
Figure 3. Test Information Curve for SDLR scale

3.9 Item and Model Fit

Item fit was assessed using the generalized item fit statistics $S - \chi^2$, which tend to control the type I error rate (Kang & Chen, 2011). Items fit indices were computed, items with a probability less than 0.05 were considered as misfit items. The results showed that the GRM better fit the data than the partial credit model; 16 of 40 items did not fit the GRM model (SM2, SM3, SM4, SM10, SM13, LD1, LD3, LD6, LD9, LD11, LD12, SC2, SC3, SC5,and SC13).

The overall goodness of fit for GRM was once examined for the 40 items scale and additionally for the 24 items scale after eliminating misfit and low discriminated items, using the M_2 statistic (Maydeu-Olivares & Joe, 2006). GRM fit statistic for the 40 SDLR scale was (M_2=17937.95, Prob. = 0.0001, RMSEA= 0.21),while these values for 24 SDLR scale were(M_2=114.30, Prob. = 0.0001, RMSEA= 0.00). This means that the 16 misfit items had significant violations for GRM. Marginal reliability for the 40 SDLR scale was 0.93, and 0.90 for the 24 SDLR scale, which means that the scale was still reliable after eliminating 16 items.

4. Discussion

The main goal for the current study is to evaluate the psychometric properties of SDLR scale and its availability as a measure of SDLR using item response theory models, while most researchers examined its properties using CTT techniques and factor analysis.

CTT is a traditional approach to improve measurements. The theory is mainly concerned with observed scores, true scores, and the error score. The theory assumes relatively simple assumptions, but its results are sample specific, rendering it insufficient (Budgell, Raju, &Quartetti, 1995; Hambleton, Swaminathan, & Rogers, 1991; Hulin, Drasgow, & Pearson, 1983).

One of IRT's most important properties is that item characteristics are independent upon a person's characteristics (Baker, 2001). Recently attention has been devoted to IRT models capable of analyzing rated data using either ordinal, categorical, or nominal scales, which make it applicable to any type of psychological assessment instrument. For the advantages discussed here, the current study analyzed the data of SDLR scale using Samejima's graded response model.

The results of discrimination parameters for each individual item showed that some items discriminate better than others between respondents, such as SC1, SC6, and SC15; the rest of the scale items discriminate well with the existence of diversity across items. This result is close to the results obtained by Fisher et al. (2001). The three items had relatively high values of corrected item total correlation, which is considered as an indicator of item discrimination (Cappelleri, Lundy, & Hays, 2015).

The results of item threshold parameter showed that most of the scale items tend to measure moderate and low levels of self directed learning except items SM4, LD5, and LD6, where it was suitable to measure the high ability levels of self directed learning.

Results of item fit statistic $S - \chi^2$ showed that only 23 items of the SDLR scale fitted the model well. For the overall goodness of fit of the GRM: M_2 showed that these 23 items fit the model well, while the overall 40 items scale had an unacceptable value of M_2, which means that these violations were significant.

The results of the current study revealed important information about the SDLR scale not previously identified. In general, the SDLR scale could be considered a good scale with some features that require improvement. The evaluated version of the 23 SDLR scale could be considered as a brief scale for SDLR, which is valid and accurate for identifying moderate and low levels of self directed learning readiness. One recommendation is that some items could be added for the brief SDLR scale with high discrimination parameters and high threshold values to measure high levels of self directed readiness.

5. Conclusion

The current study considered as the evidence for the usefulness of using IRT analysis to provide important additional psychometric information, which will impact the quality of the instruments.

Competing Interesting Statement

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Attorresi, H., Abal, F., Galibert, M., Lozzia, G., & Aguerri, M. (2011). Aplicación del modelo de respuesta graduada a una escala de voluntad de trabajo. *Interdisciplinaria, 28*(2), 231-244.

AlHassoun, Kh. (2017). *Self Directed Learning Readiness and its Relationship with Emotional Intelligence and Academic Achievement among Secondary Female students in Buraidah* (Unpublished Master Thesis). Qassim University, Qassim, KSA.

Baker, F. (2001). *The Basics of Item Response Theory* (2nd ed). USA: ERIC Clearinghouse on Assessment & Evaluation.

Budgell, G. R., Raju, N. S., & Quartetti, D. A. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement, 19*, 309-321. https://doi.org/10.1177/014662169501900401

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO for Windows* [Computer software]. Lincolnwood, IL: Scientific Software International.

Candy, P. C. (1991). *Self-Direction for Lifelong Learning: A Comprehensive Guide to Theory and Practice*. Jossey-Bass Publishers, San Francisco, CA

Cappelleri, J., Lundy, J., & Hays, R. (2015). Overview of Classical Test Theory and Item Response Theory for Quantitative Assessment of Items in Developing Patient-Reported Outcome Measures. *Clin Ther, 36*(5), 648-662. https://doi.org/10.1016/j.clinthera.2014.04.006

De Mars, C. (2010). *Item Response Theory*. New York, NY: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195377033.001.0001

Fisher, M., King, J., & Tague, G. (2001). Development of a self-directed learning readiness scale for nursing education. Nurse Educ.*Today, 21*, 516-525. https://doi.org/10.1054/nedt.2001.0589

Fisher, M., & King, J (2010). The self-directed learning readiness scale for nursing education revisited: a confirmatory factor analysis. *Nurse Educ. Today, 30*, 44-48. https://doi.org/10.1016/j.nedt.2009.05.020

Guglielmino, L. M. (1977). Development of the Self-Directed Learning Readiness Scale (Unpublished doctoral dissertation). University of Georgia. *Dissertation Abstracts International, 38*(11a), 6467.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hendry, G., & Ginns, P. (2009). Readiness for self-directed learning: validation of a new scale with medical students. *Med. Teac, 31*, 918-920. https://doi.org/10.1080/01421590802520899

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.

Kang, T., & Chen, T. T. (2011). Performance of the generalized S-X2 item fit index for the graded response model. *Asia Pacific Educ. Rev, 12*(1), 89-96. https://doi.org/10.1007/s12564-010-9082-4

Knowles, M. S.(1975). *Self-Directed Learning: A Guide for Learners and Teachers*. Association Press,New York, NY.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*, 713-732. https://doi.org/10.1007/s11336-005-1295-9

Nguyen, T., Han, H., Kim, M., & Chan, K. (2014). An Introduction to Item Response Theory for Patient-Reported Outcome Measurement. *Patient, 7*(1), 23-35. https://doi.org/10.1007/s40271-013-0041-0

O'Kell, S. P. (1988). A study of the relationships between learning style, readiness for self-directed learning and teaching preference of learner nurses in one health district. *Nurse Education Today, 8*, 197-204. https://doi.org/10.1016/0260-6917(88)90149-9

Reckase, M. D. (1979). Unifactor Latent Trait Models Applied to Multifactors Tests: Results and Implitations. *Journal of Educational Statistics, 4*, 207- 230. https://doi.org/10.2307/1164671

Samejima, F. (1969). Estimation of latent trait ability using a response pattern of gradedscores. *Psychometric Monograph, 18*. https://doi.org/10.1007/BF03372160

Steinberg, L., & Thissen, D. (1995). Item response theory in personality research. In P. E. Shrout, & S. T. Fiske (eds.), *Personality research, methods, and theory, a festschrift honoring*. Donald W. Fiske. Hillside, NJ: Lawrence Erlbaum Associates.

Torabi, N., Abdollahi, B., Aslani, G., & Bahrami, A. (2013). A Validation of Self-Directed Learning Readiness Scale among Preliminary School Teachers in Esfahan. *Pcocedia-Social and Behavioral Sciences, 83*, 995-999. https://doi.org/10.1016/j.sbspro.2013.06.185

Wiley, K. (1983). Effects of a self-directed learning project and preference for structure on self-directed learning readiness. *Nursing Research, 32*(3), 181-185. https://doi.org/10.1097/00006199-198305000-00011

Williams, B., & Brown, T. (2013). A confirmatory factor analysis of the Self Directed Learning Readiness Scale. *Nursing and Health Sciences, 15*, 430-436. https://doi.org/10.1111/nhs.12046

Copyrights