

Evaluation of an Arabic version of Children's Self-report Social Skills Scale
(CS⁴) based on Item Response Theory

November 30, 2010

Dr. Mohamed Habashy Hussein
Associate professor of Educational Psychology
College of Education
University of Alexandria, Egypt

Author's note

Correspondence concerning this work can be directed to the author at: Educational Psychology Department, College of Education, Alexandria University, Alexandria, Egypt; email: mhhussain@kku.edu.sa

Evaluation of an Arabic version of Children's Self-report Social Skills Scale (CS⁴) based on Item Response Theory

Abstract

The present study examined the psychometric properties of the Arabic version of Children's Self-report Social Skills Scale (CS⁴) using a generalized partial credit model (GPCM). Data from 722 primary school children (401 boys and 321 girls) responses, in Egypt, were analyzed using GPCM. The results indicated that the 21 items are able to discriminate among the levels of children's social skills. The item and test information functions indicated that the three subscales were more informative for low and medium levels of the social skills. Items 6, 13, and 17 showed a poor fit to the GPCM, and these items could be removed to improve the psychometric properties of the CS⁴. Overall, current findings suggest that evaluation of social skills among Egyptian elementary school children using the 21-item of the CS⁴ may usefully choose to focus on items that performed well in these IRT analyses.

Evaluation of an Arabic version of Children's Self-report Social Skills Scale (CS⁴) based on Item Response Theory

Dr. Mohamed Habashy Hussein
Associate professor of Educational Psychology
College of Education
University of Alexandria, Egypt

Introduction

Today, Schools are held accountable for improving social competencies as well as raising academic achievement to prepare students for successful adulthood and citizenship. Therefore, one of the purposes of elementary education is to develop social skills among children because social skills are just as important as academics (Rashid, 2010).

Social skills are defined as communicating, understanding other people, acting according to social environments, making friends, displaying acceptable behaviors, expressing oneself, dealing with problems and establishing a good relationship with the environment (Samanci, 2010). Social skills also defined as understanding both one's own and other individuals' feelings, thoughts and behaviors related to various interactions, and behaving according to that understanding (Sahin, 2010).

Akkok (as cited in Sahin, 2010) has classified social skills in six groups. These are: (1) skills for initiating the relationship and continuing; (2) skills for teamwork; (3) feeling-oriented skills; (4) skills for coping with aggressive behaviors; (4) skills for coping with stressful situations, and (5) skills for problem solving and planning.

Social skills are a central part of learning, playing, and behaviors such as sharing, helping, initiating communications, and requesting help from another person, and giving compliments.

Children who lack important social skills often are rejected by their peers, have trouble interacting with their teachers and families, and have emotional difficulties. Furthermore, social skill deficits are related to poor academic performance, and frequently associated with children exhibiting externalizing disorders such as delinquency and conduct disorder, as well as those with internalizing disorder like depression and anxiety (Warnes, Sheridan, Geske, & Warnes, 2005). Impairments in social skills are related to a broad range of problems including juvenile delinquency, ADHD, developmental disabilities, social isolation and withdrawal, aggression and antisocial behavior, mental health problems, and dropping out of schools (Matson & Wilkins, 2009).

By contrast, children whose levels of social skills are high represent kids who are able to adjust to their environment, succeed in avoiding conflict, and maintain good communications with others (Cummings, Kaminski, & Merrell, 2008). Children who develop adequate social skills tend to exhibit fewer problems with adults and peers, and better adjustment in society (Shahim, 2004).

Several theories highlighted the interdependency of emotional and social competence (e.g., Denham, 2007; Rose-Krasnor, 1997). Social interactions and relationships are guided even defined by the emotional transactions within them.

On the other hand, Emotional competence is crucial to children's ability to interact and form relationships with others. Knappmeyer, Thornton, and Bulthoff (2003) found a link between a decreased ability to recognize emotion and social dysfunction.

Socially competent children should be more effective in recognizing

Social Skills Scale

emotions in others and in themselves, regulating their own emotional experience, and sympathizing with the emotions of the peers. Conversely, children who are emotionally troubled lack the scale to establish and sustain successful relationships with their peer and teachers (Yukay-Yuksel, 2009).

Social Skill Assessment

Identification and treatment of children with low social skills are important tasks for educators, psychologists, and mental health professionals. Screening and assessment are essential foundations for effective intervention in social-behavioral problem of children and youth. Without the careful Identification, classification, and selection that should be a part of a good assessment, social behavior intervention is likely to be haphazard and disorganized at best and ineffective at worst (Merrell, 2001).

Elliott and Gresham (1987) provided a heuristic framework for conceptualization the assessment of children's social skills that include: (a) teacher, parent, and student ratings, (b) teacher, parent, and student interview, (c) observation, (d) behavioral role playing, and (e) sociometric techniques. Merrell (2001) mentioned that there are six primary methods of gathering assessment information about children's social skills: behavioral observation, behavioral rating scales, interviewing, self-report instruments, projective expressive techniques, and sociometric techniques. Matson and Wilkins (2009) classified the methods of assessing social skills into two primary approaches: Standardized role-play scenes and tests of social behavior with a list of items that can be scored in Likert's format.

Although all of these aforementioned methods have some advantages and disadvantages, self-report inventories are one of the most straightforward ways of measuring social skills (Elliott, Busse, & Gresham, 1993). Advantages of child self-report measures include the following: (a) The instruments are generally inexpensive in terms of administration time, and can be easily administrated in a wide variety of settings such as schools (Beitchman & Corradini, 1988), (b) assessing feelings and tendencies over a wide range of unobservable social behaviors and situations (Segrin, 2000), and (c) meaningful information –the child's perception and cognitions- is provided that is not otherwise accessible to other reporters.

In USA a number of scales to measure children's skills have been developed. The Children's Self-Report Social Skills CS⁴; (Danielson & Phelps, 2003) is one of these scales and is frequently used in the USA. However, this scale is not well known in Arabic countries, and it has not been evaluated with an Arabic population. Furthermore, there have been a few scales to measure children's social skills developed in Arab countries, in general, and in Egypt, in particular.

The Arabic version of Riggio's Social skill Inventory (SSI) which was translated by Samadoni (1991) is one of the most common scales that measures social skills. The SSI is a 90-item instrument in which subjects rate themselves on both positive and negative social behaviors. The scale applies to subjects with 14 years and over reading at or above the eighth grade level. Three disadvantages are associated with the SSI: The psychometric properties of the instrument were estimated through Classical Test Theory (CTT), which is a sample dependent; the instrument is too long to use as screening instrument; and the original version of the instrument is over 20 years old.

Item Response Theory versus Classical Test Theory

Although the CS⁴ has indeed been a promising instrument, it is noteworthy

that extant work has principally evaluated this instrument using CTT methodologies (Danielson and Phelps, 2003; Gençdoğan, 2008). CTT has a number of limitations, including item-dependent estimates, and unconditional standard error of measurement (see Magno, 2009; Rouse, Finger, and Butcher, 1999).

To address these types of concern, Item Response Theory (IRT) has proven valuable advantages over CTT. IRT offers some unique advantages included: (a) detailed description of the performance of individual items, (b) indices of item and scale precision that are free to vary across the full range of possible scores, (c) assessments of item and test level bias with respect to demographic subgroups, (d) measures of response-profile quality, and (e) computer-adaptive testing, which can dramatically reduce testing time (Hall, Hidalgo, Tomas-Sabado, 2007; Reeve, Hays, and Chang, 2007).

Indeed, IRT has been successfully employed to refine the assessment of a number of psychological instruments (e.g., Becker, Schwartz, Saris-Baglama, Kosiniski, and Bjorner, 2007; Cooke, Kosson, and Michie, 2001; Gomez, Hidalgo, and Tomas-Sabado, 2007; Hall, Reise, and Haviland, 2007; Takegami et al., 2009; Zvolensky, Strong, Vujanovic, and Marshall, 2009).

With this background, the aims of the present study were to evaluate the psychometric properties of CS⁴ by using both the CTT and the IRT models, and to underline the differences and similarities between the two models.

Method

Participants

Participants were 722 from 5th and 6th grade students attended public elementary schools in Alexandria, Egypt. Of the 722 students, 401 were boys, 321 girls, with 403 children in the 5th grade (266 boys and 177 girls), and 319 in the 6th grade (175 boys and 144 girls). The mean and standard deviation of the male and female participants' age were (11.52 and 0.652), and (11.40, and 0.700) respectively.

Measures

Children's Social Skills Scale CS⁴: The CS⁴ (Danielson and Phelps, 2003) is a 21-item measure in which children are asked to rate own social behavior on a 5-point Likert-type scale (1=never, 2=hardly ever, 3=sometimes, 4=most of the time, and 5=always). For the seven items that are framed to measure poor skills (e.g., speaking too loudly), points awarded to a response are reverse scored. Points are added to obtain a total score, with high scores representing greater social skills. The range of the possible scores is 21 to 105.

Emotion Awareness Questionnaire for Children Revised The EAQC-R (Rieffe, Oosterveld, Miers, Terwogt, & Ly, 2008), a modification version of EAQC (Rie et al., 2007). The EAQC-R aims to identify how children and adolescents feel and think about their feelings. It was designed with a six-factor structure describing six aspects of emotional functioning: (1) Differentiating Emotions, (2) Verbal Sharing, (3) Not-Hiding Emotions, (4) Bodily Awareness of Emotions, (5) Attending to Others' Emotions, and (6) Analysis of Emotions. Respondents were asked to rate the degree to which each item was true about them on a three-point scale (1=not true, 2=sometimes true, 3=often true). Reliability coefficients of the six subscales of the English version varied between 0.63 and 0.68. Similarly, the six subscales of the Arabic version ranged from 0.70 to 0.75.

Translation Procedure

The original English version of the CS⁴ was first translated into the

Arabic language by the author of the present study. The initial translation was checked by a psychologist and English language expert. Based on their comments wording adjustments were made. The Arabic version was then back translated into English by a bilingually fluent researcher who had not seen the original English version. The two versions were compared, and minor wording adjustments were made based on this step.

Item response Theory Model

When item responses are coded into more than two ordered categories, polytomous ordered response models are appropriate in order to know the test psychometric properties. The graded response model of Samejima (1969) and Muraki (1992) generalized partial credit model (GPCM) are item response models that can be applied to Likert-type scaled items. However, the GPCM offers great flexibility and potential benefits when working with polytomous items (Fox, 1999; Masters, 1988).

On the other hand, Edelen and Reeve (2007) believe that the choice between these two models is somewhat arbitrary, as they generally produce nearly identical results, albeit with slightly different parameterizations, and choosing one of these two models over the other tends to be primarily a result of personal preference and familiarity with software (PARSCALE is set up to estimate the GPCM more easily, whereas MULTILOG favors the GRM).

In the current study, GPCM was used to calibrate the CS⁴. The GPCM estimates two main parameters: the slope or discrimination parameter, and the threshold or step difficulty parameters which are associated with the transition from one category to the next, the number of step difficulty parameters equals to the number of categories minus one (Embretson & Reise, 2000).

Results

Results of Classical Test Theory

As shown in Table 1, the overall mean of the girls is slightly higher than the corresponding mean of boys. On the other hand, the means of boys and girls on Social Rules and Likeability are very close, whereas the mean of Social Ingenuousness are smaller for girls than boys indicating that boys are more Social Ingenuousness than girls. The variability of the three scales is very similar across boys and girls. The reliability coefficients of the three subscales and the total scores are significant and fell above the criterion 0.70 indicating that the scale and the three subscales are reliable for boys and girls. The items of the CS⁴ were also evaluated through calculation of correlation coefficients between the score on each item and the total score. For all the items, except item 15, the correlation coefficients were significant and greater than 0.30, which met Danielson and Phelps' (2003) criterion for good item-scale correlation. Only item 15 ($r=0.16$) failed to meet this criterion, although this correlation was still statistically significant.

Evaluation of the Model Assumption

The assumptions of unidimensionality and local independence were evaluated. The two concepts are related. Therefore, a data set is unidimensional when the item response is locally independent based on a single latent trait. Because of the five-point ordinal scale of the items, a polychoric correlation, which measures the linear relationship between two observed discrete variables, was computed using PRELIS 2.5.4.

Because the CS⁴ was used for the first time in the Arabic culture, exploratory factor analysis was used, and the number of factor was not fixed. Three criteria were used to determine the number of factors: (1) the Eigenvalue

Table 1: Means, Standard Deviations, and Cronbach's Alpha Coefficients of CS⁴ subscales

	Means		Standard Deviation		Alpha Coefficient	
	Boys	Girls	Boys	Girls	Boys	Girls
Social rules	39.58	39.99	6.69	6.59	0.77	0.80
Likeability	15.99	15.88	3.02	2.93	0.75	0.76
Social Ingenuousness	14.75	12.83	3.75	3.57	0.74	0.72
Total Score	80.32	82.54	10.40	10.50	0.80	0.84

should be larger than one; (2) the scree test; and (3) the content of the factor.

Eigenvalues greater than one is the most common criterion but often results in too many factors. Therefore, as a second criterion, the scree test was used, a graph of the Eigenvalues for each factor in which one look for breaks in the graph. Items with an absolute factor loading smaller than 0.40 in the pattern matrix of the principal axis analysis with Promax rotation were excluded.

SPSS 18 was used to explore the factorial structure of CS⁴. As shown in Table 2, for both boys and girls, eigenvalue criterion extracted three factors, and their Eigenvalues were greater than one (boys: 3.46, 1.96, 1.43, and girls: 3.77, 2.00, 1.41). Similarly, the Scree Plot test showed three factors.

The three factors explained 31.97% of the variance for boys, and 34.17% for girls. The correlations among the three factors indicated that the subscales related to each other, but the size of the correlation coefficients, which ranged from 0.13 to 0.54, showed that the subscales measured separate constructs. The facts that all, but one item, loaded with a factor loading greater than 0.40 on one and only one CS⁴ factor provided evidence on the unidimensionality, which supported our use of IRT to psychometrically describe the scale, of each subscale.

For each subscale, the unidimensional generalized partial credit model (GPCM), which offers better understanding about how each item and subscale measure children's social skills, had been fitted using marginal maximum likelihood estimation procedures implemented in PARSCALE 4.1 (Muraki & Block, 2003).

Since, it is possible to test directly for each item the model-data fit in relation to the GPCM, a likelihood-ratio χ^2 which indicates the goodness of fit between the expected and observed response frequencies was computed. Non-significant difference between the expected and the observed frequencies indicates a good fit.

Construct Validity

In the current study, the construct validity of CS⁴ was investigated through computing the correlation coefficients between emotional awareness as measured by Emotion Awareness Questionnaire for Children Revised (Rieffe, Oosterveld, Miers, Terwogt, & Ly, 2008) and the components of the CS⁴. We hypothesized that the positive components of the CS⁴ (Social Rules and Likeability) will correlate positively with the total score The EAQC-R, whereas Social Ingenuousness as a negative component of the CS⁴ will correlate negatively with the total score on the EAQC-R.

The findings from the present study indicated that correlation coefficient between the CS⁴ and Emotion Awareness Questionnaire-Revised (EAQC-R) was statistically significant for both boys and girls. Pearson's correlation coefficient demonstrated a significant positive correlation between child's total scores on the CS⁴ and the scores on the EAQC-R for both boys ($r=0.38$, $p \leq 0.01$) and girls ($r=0.46$, $p \leq 0.01$). This result suggested that higher scores on the self-report social skills were associated with higher scores on EAQC-R.

Social Skills Scale

Table 2: Pattern Coefficients, Eigenvalues, and Variance accounted for by the items on the three components of the Children’s Self-Report Social Skills Scale for Boys and Girls

	Components					
	Social Rules		Likeability		Social Ingenuousness	
	Boys	Girls	Boys	Girls	Boys	Girls
1. I look others in the face when they talk.	0.34	0.30				
3. I say Thank you when someone does something nice for me.	0.57	0.41				
6. I take turns with others.	0.46	0.48				
9. I listen to others when they talk.	0.43	0.47				
10. I share games and toys with others.	0.44	0.56				
11. I say I am sorry when I hurt someone by accident.	0.58	0.68				
12. When I see others playing a game I would like to play, I ask if I can join them.	0.48	0.56				
14. I say I am sorry when I hurt someone on purpose.	0.42	0.52				
19. I help others when they need help.	0.42	0.65				
20. I ask others to play.	0.44	0.45				
2. Others like me and have fun with me.			0.57	0.63		
7. When I come over, others ask me to move or give them more space.			0.51	0.47		
13. I make friends easily.			0.43	0.48		
16. Others don not like me.			0.62	0.59		
18. Others ask me to play I kick or hit someone else if			0.43	0.51		
4. make me angry.					0.75	0.71
5. I am bossy.					0.74	0.69
8. I do not play fairly.					0.44	0.44
15. I walk up to others and start conversation.					0.41	0.48
17. I speak or interrupt if someone else talking.					0.74	0.45
21. I am too loud when I talk.					0.41	0.54
Eigenvalue	3.46	3.77	1.96	2.00	1.43	1.41
Variance Accounted for	16.48	17.93	9.35	9.54	6.14	6.70

The correlation coefficients between the three components of the CS⁴ (Social Rules, Likeability, and Social Ingenuousness) and the EAQC-R were examined.

The findings indicated that each component of the CS⁴ correlated statistically significant with the EAQC-R. More specifically, Social Rules and Likeability correlated positively with the EAQC-R (boys: $r=0.28$, $p \leq 0.01$, and girls: $r=0.40$,

$p \leq 0.01$), (boys: $r=0.24$, $p \leq 0.01$, and girls: $r=0.20$, $p \leq 0.01$) respectively, whereas Social Ingenuousness correlated negatively with EAQC-R (boys: $r=-0.24$, $p \leq 0.01$, and girls: $r=-0.26$, $p \leq 0.01$).

IRT parameter estimates of three subscales for boys and girls

Because the CS⁴ was found to be a three-factor model, parameters were estimated separately for each subscale. GPCM parameter estimates for the three subscales are presented in Tables 3 and 4. For GPCM, each item has a single slope (a) parameter, and a single location parameter (b), and four step difficulty parameters (d_i).

The slope parameter (comparable to discrimination) describes how well the item performs in general. Large slope value indicates that the item is good at discriminating among the different levels of the latent trait. The difficulty of an item is indicated by the location parameter. Thus, a large positive value indicates a difficult item or that few examinees respond in the higher categories. A negative value indicates an easy item or that few examinees respond in the lower categories. The location parameter functions to shift the category parameters up and down the latent trait scale.

The step difficulty parameter indicates the difficulty of the step in moving from one response option to another. For example, if an item has four response options, then there will be three steps, namely step 1 for moving from the first response option to the second option, step 2 for moving from the second option to the third option, and step3 for moving from the third option to the fourth option. Higher positive values indicate difficult steps, while low and negative values indicate easy steps. Tables 3 and 4 show the estimated item parameters for boys and girls respectively, for the three subscales of CS⁴ (Social Rules, Likeability, and Social Ingenuousness). The items are evaluated with respect to their slope parameter, their threshold parameter, and the step parameters; the latter is interpreted as the difficulty associated with a given category compared with that of other categories, or the deviation of each category threshold from the item location.

As shown in Tables 3 and 4, for the CS⁴ subscales, the values of the category parameters of step1 were negative and low relative to steps 2, 3, and 4. Although all the category parameter values for steps 3 and 4 were positive, the category parameter values for steps 1 and 2 were negative. In general, the category parameter values were monotonically increasing from step1 to step 4. These trends were consistent across the three subscales and across boys and girls.

Based on the guidelines suggested by Baker (2001), the items of the three subscales for boys and girls show moderate (0.65 to 1.34) and large (1.35 to 1.69) discrimination values. For Social Rules items, item 3 was the most discriminant item for boys, whereas, for girls, item 11 had relatively larger value than the other items. On the other hand, items 10 and 1 had relatively lower discrimination values than other items for boys and girls respectively.

For Likeability items, item 13 had relatively larger values, with the other items having about the same values for boys. For girls, item18 was the most discriminating item, followed by item 12. On the other hand, items 2 and 13 were the lowest discriminating items for boys and girls respectively.

For Social Ingenuousness items, items 5 and 4 were the most discriminating item for boys and girls respectively, whereas item 17 was the lowest discriminating item for both boys and girls.

In relation to location parameters, although the ten items of the Social Rules subscales had negative values, item 19 was the easiest item for boys and girls.

Social Skills Scale

Table 3: Generalized Partial Credit Model Item Parameter Values for Boys

items	Category				slope	location	G2	df	p
	step1	step2	step3	step4					
Social Rules									
1.	-3.08	-1.67	1.41	3.34	0.75	-0.66	37.41	33	0.27
3.	-2.06	-0.42	0.68	1.80	1.26	-1.39	17.22	11	0.10
6.	-2.53	-1.34	1.19	3.69	0.79	-1.33	37.33	23	0.03
9.	-2.49	-1.41	0.57	3.33	0.82	-1.11	30.62	21	0.08
10	-6.51	-4.79	2.70	5.59	0.68	-3.88	28.55	22	0.16
11.	-3.40	-2.04	2.64	2.80	0.88	-1.22	22.19	16	0.14
12	-1.60	0.32	0.64	1.64	0.88	-1.46	29.55	23	0.16
14.	-3.22	-1.13	1.40	4.95	0.76	-0.58	37.94	26	0.06
19.	-5.36	-2.02	2.06	3.32	0.70	-4.55	16.94	11	0.11
20	-3.17	0.92	0.95	1.30	0.78	-1.88	33.97	28	0.20
Likeability									
2.	-6.41	-4.32	4.54	6.55	0.61	0.00	28.77	26	0.32
7.	-3.41	-2.71	1.69	4.88	0.69	-2.72	28.92	23	0.21
13	-1.11	-0.64	0.16	1.54	0.88	-1.06	48.08	24	0.00
16.	-4.21	-2.71	3.03	4.55	0.72	0.11	42.21	35	0.19
18.	-2.40	-0.47	0.53	2.34	0.82	-1.47	30.51	25	0.21
Social Ingenuousness									
4.	-1.22	-0.36	-0.09	1.67	0.97	-0.12	40.27	32	0.15
5.	-0.63	-0.58	0.25	0.96	1.42	-0.38	27.99	19	0.08
8.	-8.06	-2.09	3.09	5.79	0.72	-2.41	17.84	16	0.33
15.	-2.47	-2.30	2.10	2.54	0.79	-0.97	37.02	28	0.12
17.	-6.05	-4.51	4.52	8.04	0.61	2.79	59.02	36	0.01
21.	-6.22	-3.82	3.59	6.44	0.72	0.00	34.94	28	0.17

For the Likeability items, item 16 was the most difficult item for girls and boys, whereas items 7 and 2 were the easiest items for boys and girls respectively. Finally, for Social Ingenuousness items, item 17 was the most difficult item for boys and girls, whereas item 8 was the easiest items for boys and girls.

Social Skills Scale

Table 4: Generalized Partial Credit Model Item Parameter Values for Girls

items	Category				slope	location	G2	df	p
	step1	step2	step3	step4					
Social Rules									
1.	-5.64	-1.30	2.23	4.53	0.69	-0.98	29.87	26	0.27
3.	-3.09	-2.53	1.50	4.11	0.89	-1.98	7.01	3	0.07
6.	-4.35	-2.95	3.38	6.91	0.74	0.00	54.91	26	0.00
9.	-2.92	-0.58	0.99	2.51	0.88	-0.84	19.24	17	0.13
10	-3.76	-0.59	1.97	2.38	0.85	-1.41	21.28	17	0.21
11.	-0.82	-0.21	0.03	1.01	1.55	-1.39	18.57	11	0.07
12	-8.69	-7.01	7.21	8.84	0.70	-0.11	37.22	26	0.07
14.	-3.67	-0.97	1.56	3.38	0.76	-0.84	29.21	23	0.17
19.	-1.47	-0.10	0.68	0.90	1.03	-1.62	16.09	12	0.19
20	-2.99	-0.38	0.82	2.55	0.81	-1.41	26.59	22	0.23
Likeability									
2.	-1.06	-0.31	0.22	1.16	1.01	-1.20	21.33	17	0.21
7.	-6.36	-3.76	2.50	7.61	0.70	0.60	17.52	15	0.29
13	-6.21	-3.87	3.39	7.68	0.60	0.00	52.71	27	0.00
16.	-5.73	-4.64	3.64	6.21	0.62	2.83	28.21	24	0.25
18.	-1.01	-0.50	-0.10	1.61	1.30	-1.04	21.78	19	0.30
Social Ingenuousness									
4.	-0.56	-0.34	0.36	0.54	1.26	-0.69	18.45	14	0.19
5.	-1.86	-0.88	0.52	2.22	0.90	-1.31	16.13	14	0.31
8.	-2.24	-1.23	1.24	3.32	0.80	-2.11	18.72	13	0.13
15.	-2.26	-0.75	0.87	1.98	0.79	-0.81	24.35	20	0.23
17.	-6.16	-4.04	3.52	7.52	0.72	2.79	50.59	27	0.00
21.	-0.67	-0.34	-0.18	1.19	0.80	-2.08	26.85	18	0.08

Item-Fit Statistics

There is not widely accepted goodness-of-fit statistic for polytomous IRT models. Several strategies can be used (Embretson & Reise, 2000). In the current study, the residuals by item and response category were calculated to assess the fit. Moreover, the log-likelihood item-fit Chi-square statistic (G2)

was calculated using the PARSCALE computer program. Although this test is highly sensitive to sample size (Gomez et al., 2007), and probably should not be treated as a solid decision-making tool. As well, G2 statistic is sensitive to the number of intervals into which the ability continuum is divided, with a high number of intervals, the values of this statistic maybe artificially high (Muraki, 1992).

Tables 3 and 4 display the G2 statistics for each item, their associated degree of freedom (df), and their probability value (p). The results were consistent across boys and girls, the same three items, one from each subscale, present asset value of chip-square, which were significant($p \leq 0.05$), indicating their poor fit with the model for boys and girls. For Social Rules items, item 6 shows a poor fit, whereas item 13 shows a poor fit for Likeability, item17 shows a poor fit for Social Ingenuousness.

For boys, the overall goodness-of-fit statistics of the three subscales (Social Rules, Likeability, Social Ingenuousness) were $G^2=(305.72, 535.12, 237.08)$, with $df=(214, 386, 159)$, and $p \leq 0.001$ respectively. Similarly, for girls they were $G^2 = (305.72, 535.12, 237.08)$, with $df = (214, 386, 159)$, and $p \leq 0.001$ indicating a not good overall fit to the data. Moreover, Chi-Square Ratio Test (G^2/df): A significant chi-square value relative to the degrees of freedom indicates that the observed and estimated matrices differ. Statistical significance indicates that this difference is due to sampling variation. A chi square/degrees of freedom ratio value of 2-3 will be interpreted as suggesting a plausible model (Carmines & McIver, 1981).The chi-square/degrees of freedom statistics for the three subscales for boys (1.43, 1.39, 1.49) and girls (1.51, 1.29, 1.43) were ≤ 2 indicating a good fit.

Item Information and Test Information Functions

The Item Information Function (IIF) shows trait levels where the item has more precision and reliability; that is, it indicates in which trait levels the item is most informative. The notion of reliability from classical test theory (CTT) is analogous to item response theory information (IRT). In CTT, however, reliability is summarized in a single coefficient that represents the average precision across all examinees. In IRT, information function plots show the varying precision of the ability estimate across the trait continuum.

IIF is mainly based on the value of the item discrimination, as the item slope increases, the IIF of this item increases items. Therefore, the three most discriminative items were the most informative items. For boys, items 3, 13, and 5 were the most informative items, whereas items 10, 2, and 17 were the least informative items for Social Rules, Likeability, and Social Ingenuousness respectively. For girls, items11, 18, and 4 were the most informative items, whereas items 6, 13, and 17 were the least informative items for Social Rules, Likeability, and Social Ingenuousness respectively. The most informative items are the most discriminating (having highest slope parameter values), whereas the least informative items are the least discriminating (having lowest slope parameter values).

The Test Information Function (TIF) is the sum of value of IIF of subscale items. While the number of items of the subscale increases the value of the TIF of the subscale or the subscale increases. In the present study, TIF was calculated for each subscale. These graphical representations show the trait levels where the tests are most discriminating. They also provide an overall standard error of measurement (SEM) given by the inverse square root of the TIF at each trait level. In other words, the TIF indicates for each level of the latent trait the amount of precision expected when estimating a subject trait. For example, if a scale designed to measure the social skills has a low SEM at the upper end of the trait continuum and a high SEM at the lower end, then the clinician using the scale is aware that high scores on the scale are fairly accurate estimates of the subject's social skills.

Social Skills Scale

Therefore, it is appropriate to use this instrument to detect subjects with high trait levels. In contrast, when this questionnaire is used with subjects with low trait levels, it will provide inaccurate estimates of these levels. That is, the instrument will not be able to detect differences between subjects with medium and below medium trait levels. Thus, the TIF serves as a guide in deciding when and for what purpose a given instrument may be best be applied.

For both boys and girls, the TIF of each subscales shows that the Social Rules subscale (consists of the largest number of items) provides the most amount of information along the trait continuum. However, the subscale of Social Rules is most informative at medium or low levels of Social Rules. The Social Ingenuousness was the second informative subscales for boys and girls. Similarly, the Social Ingenuousness provides the most amount of information at low levels of Social Ingenuousness. Finally, Likeability was the least informative subscale, although it was higher for girls than boys. The overall conclusion is that the three subscales are more accurate in estimating the low trait levels of social skills, whereas, they are inaccurate in estimating the high levels of social skills for boys and girls.

Discussion

The aim of present study was to utilize GPCM to evaluate the psychometric properties of the Arabic version of CS⁴ for boys and girls. TO compare the results from the present study to the results from the previous ones (Danielson and Phelps, 2003; Gençdoğan, 2008), the psychometric properties of the CS⁴ were also evaluated based on the classical test theory.

The psychometric properties of CS⁴ based on CTT in the present study were very comparable with the results from the English version (Danielson & Phelps, 2003), and the Turkish version (Gençdoğan, 2008). In present study, the internal consistencies of the CS⁴ total score and the three subscales were acceptable, and they were close to the ranges that were obtained from the original English version (Danielson & Phelps, 2003). The exploratory factor analysis revealed that CS⁴ consists of three factors. This finding is consistent with the Danielson and Phelps' three-factor structure model. These findings provide a preliminary indication of the stability of the CS⁴ across languages and countries. Such stability suggests that the CS⁴ accurately captures the structure of the children's social skills regardless of the cultural differences.

On the other side, the main difference between the results from the present study and the previous ones was that two items from the Social Rules scale loaded on Social Ingenuousness scale. Items 4 ("I kick or hit someone else if they make me angry"), and 5 ("I am bossy") loaded differently from the English and Turkish versions. The inspection of the content of the two items indicated that they are more related with Social Ingenuousness subscale than Social Rules subscale.

The CS⁴ showed an adequately fit to the generalized partial credit model. Thus, for boys and girls and across the three subscales, only three out of the 21 items presented a poor fit to the model. Embretson and Reise (2000) noted that, in general, some sources of a poor item-fit may be: (a) multidimensionality, (b) a failure to estimate enough item parameters(e.g., when polytomous Rasch model is fitted to data that had slope parameter variations), (c) nonmonotonicity of item-trait relations, or (d) poor item construction.

In this study, the poor fit of these three items is more likely related to issues of nonmonotonicity and wording. The Item Characteristic Curves (ICCs) of two (13 and 17)of three items are flat. Thus, as the trait increases the probability of endorsing the options are constant. The two items (item 13 "I make friends easily", and item 17 "I speak or interrupt, if someone else is talking") imply high social desirability for Arabian children. On the other hand, item 6 ("I take a turn with others") are worded in a way that is perhaps abstract and ambiguous, and may be difficult to interpret for some children. Thus, poor fit items should be revised in content and may be excluded from the test, if they do not lead to a narrow definition of the trait being measured.

In relation to slope, both boys and girls had acceptable values for all the Social Rules, Likeability, and Social Ingenuousness items. Thus all the CS⁴ items were generally good for discriminating their respective traits for boys and girls. Despite this, there was

Social Skills Scale

notable variability in discrimination abilities across the items within scales, and across boys and girls.

The present study found that for all items for boys and girls, the category parameter values for steps 1 and 2 were smaller than the values of steps 3, and 4. Thus findings suggest that for all items, moving from endorsing response options 0 to 1 and 1 to 2 is more likely than moving from endorsing response 2 to 3 or from 3 to 4. This study also found that the location parameter values for eighteen items out of the twenty one items were negative, thereby indicating that these items are easy. This implies that endorsement of higher ratings of the items would require that small amount of the relevant traits (Social Rules, Likeability, and Social Ingenuousness items) be present.

The findings from this study indicated that the three subscales of the CS⁴ had peaked information curves. Consequently, their precision differs markedly within their respective trait. In general, the three subscales of the CS⁴ tend to perform best for children at low and moderate levels of social skills. This may be exactly what one desire from social skill measure; however, this distinction could be made with fewer items on each subscale. Computing Item Information Curves, as done in this study, would allow researchers to identify exactly how many items per scale are needed to achieve a given level of precision within specific trait range.

In conclusion, this study has shown that the use of IRT procedures can provide valuable additional psychometric information over classical test theory. It is well documented that good CTT based psychometric properties for a measure do not necessarily mean that it would have good IRT based psychometric properties. This has to be demonstrated using IRT procedures. This study has also demonstrated how IRT can be used to revise existing measures. It is hoped that this study has shown the value of using IRT to evaluate the psychometric properties of measures and for test development and revision, and that it will encourage other researchers to use IRT approaches for similar purposes.

References

- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). University of Maryland, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Becker, J., Schwartz, C., Saris-Baglama, R. N., Kosiniski, M. & Bjorner, J. B. (2007). Using item response theory (IRT) for developing and evaluating the pain impact questionnaire (PIQ-6TM). *Pain Medicine*, 8(S3), S129–S144.
- Beitchman, J. H. & Corradini, A. (1988). Self-report measures for use with children: A review and comment. *Journal of Clinical Psychology*, 44(4), 477–490.
- Bell-Dolan, D. J. & Allan, W. D. (1998). Assessing elementary school children's social skills: Evaluation of the parent version of the Matson evaluation of social skills with youngsters. *Psychological Assessment*, 10(2), 140–148.
- Buhrmester, D., Furman, W., Wittenberg, M.T. & Reis, H.T. (1988). Five domains of interpersonal competence in peer relationships. *Journal of Personality and Social Psychology*, 55(6), 991–1008.
- Carmines, E. & McIver, J. (1981). Analyzing models with unobserved variables: analysis of covariance structures. In G. Bohrnstedt & E. Birgatta (Eds.). *Social measurement: current issues* (pp. 65–115). Beverly Hills, CA: Sage.
- Cooke, D. J., Kosson, D. S. & Michie, C. (2001). Psychopathy and ethnicity: structural, item and test generalizability of the psychopathy checklist-revised (PCL-R) in Caucasian and African American participants. *Psychological Assessment*, 13(4), 531–542.
- Cummings, K. D., Kaminski, R. A. & Merrell, K. W. (2008). Advances in the assessment of social competence: findings from preliminary investigation of a general outcome measure for social behavior. *Psychology in Schools*, 45(10), 930–943.
- Danielson, C. K. & Phelps, C. R. (2003). The assessment of children's social skills through self-report: a potential screening instrument for classroom use. *Measurement and Evaluation in Counseling and Development*, 35, 218–229.
- Denham, S.A. (2007). Dealing with feelings: How children negotiate the worlds of emotions and social relationships. *Cognition, Brain and Behavior*, 11(1), 1-48.
- Edelen, M.O. & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*, 16, 5–18.
- Elliott, S. N., Busse, R. T. & Gresham, F. M. (1993). Behavior rating scales: issues of use and development. *School Psychology review*, 22(2), 313–321.
- Elliott, S. N. & Gresham, F. M. (1987). Children's social skills: assessment and

Social Skills Scale

- classification practices. *Journal of Counseling and Development*, 66, 96–99.
- Elliott, S. N., Malecki, C. K. & Demaray, M. K. (2001). New directions in social skills assessment and intervention for elementary and middle school students. *Exceptionality*, (1&2), 19–32.
- Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists* (3rd ed.). Mahwah: Lawrence Erlbaum Associates.
- Fergusson, D. M., Woodward, L. J. & Horwood, L. J. (1999). Childhood peer relationship problems and young people's involvement with deviant peers in adolescence. *Journal of Abnormal Child Psychology*, 27, 357–370.
- Fox, C. (1999). An introduction to the partial credit model for developing nursing assessment. *Journal of Nursing Education*, 38(8), 340–346.
- Gençdoğan, B. (2008). Psychometric properties of the Turkish version of the children's self-report social skills scale. *Social Behavior and Personality*, 36(4), 955–964.
- Gomez, J., Hidalgo, M. D. & Tomas-Sabado, J. (2007). Using polytomous item response models to assess death anxiety. *Nursing Research*, 56(2), 89–96.
- Hall, T. W., Reise, S. P. & Haviland, M. G. (2007). An item response theory analysis of the spiritual assessment inventory. *The International Journal for the Psychology of Religion*, 17(2), 157–178.
- Horn, M. L. V., Atkins-Burnett, S., Karlin, E., Ramey, S. L. & Snyder, S. (1988). Parent rating of children's social skills: longitudinal psychometric analysis of the social skills rating system. *School Psychology Quarterly*, 22(2), 162–199.
- Hussein, M. H. (2009). The psychometric properties of the Arabic version of emotional awareness questionnaire. *Egyptian Journal of Psychological Studies*, 16, 15–38.
- Knappmeyer, B., Thornton, I. M., & Bulthoff, H. H. (2003). The use of facial motion and facial form during the processing of identity. *Vision Research*, 43, 1921–1936.
- Ladd, G. W. & Troop-Gordon, W. (2000). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development*, 74(5), 1344–1367.
- Magno, C. (2009). Demonstrating the differences between classical test theory and item response theory using derived test data. *The International Journal of Education and Psychological Assessment*, 1(1), 1–11.
- Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education*, 1(4), 279–397.
- Matson, J. L. & Wilkins, J. (2009). Psychometric testing methods for children's social skills. *Research Developmental Disabilities*, 30, 249–274.
- Merrell, K. W. (2001). Assessment of children's social skills recent development, best practices, and new directions. *Exceptionality*, 9(1&2), 3–18.
- Muraki, E. (1992). A generalized partial credit model: application of an em

- algorithm. *Applied Psychological Measurement*, 16, 159–176.
- Muraki, E. & Block, R.D. (2003). PARSCALE (Version 4.1) [Computer software]. Chicago, IL: Scientific Software, Inc.
- Rashid, T. (2010). Development of social skills among children at elementary level. *Bulletin of Education and research*, 32(1), 69-78.
- Reeve, B. B., Hays, R. D. & Chang, C. (2007). Applying item response theory to enhance health outcomes. *Qual Life Res*, 16, 1–3.
- Reise, S.P. & Henson, J. M. (2000). Computerization and Adaptive Administration of the NEO PI-R. *Assessment*, 7(4), 347–364.
- Rieffe, C., Oosterveld, P., Miers, A.C., Terwogt, M.M. & Ly, V. (2008). Emotion awareness and internalising symptoms in children and adolescents: the emotion awareness questionnaire revised. *Personality and Individual Differences*, 45, 756–761.
- Rieffe, C., Terwogt, M. M., Petrides, K. V., Cowan, R., Miers, A. C. & Tolland, A. (2007). Psychometric properties of the emotion awareness questionnaire for children. *Personality and Individual Differences*, 43, 95–105.
- Riggio, R. E. (1986). Assessment of basic social skills. *Journal of Personality and Social Psychology*, 51(3), 649–660.
- Rose-Krasnor, L. (1997). The nature of social competence: A theoretical review. *Social Development*, 6(1), 111-135.
- Rouse, S. V., Finger, M. S. & Butcher, J. N. (1999). Advances in clinical personality measurement: an item response theory analysis of the mmpi-2 psy-5 scales. *Journal of Personality Assessment*, 72(2), 282–307.
- Sahin, C. (2010). An evaluation of teachers' perception of teaching social skills to fifth grade students within the scope of social studies lessons. *International Journal of Progressive Education*, 6(1), 28-46.
- Samadoni, E. E. (1991). *Social skills scales*. Cairo: The Anglo Egyptian - bookshop.
- Samanci, O. (2010). Teacher view of social skills development in primary school students. *Education*, 131(1), 147-157.
- Samejima, F. (1969). Estimation of ability using a response to pattern of graded scores. *Psychometric Monograph*, (NO 17).
- Segrin, C. (1998). The impact of assessment procedures on the relationship between paper and pencil and behavioral indicators of social skills. *Journal of Nonverbal Behavior*, 22(4), 229–251.
- Segrin, C. (2000). Social skills deficits associated with depression. *Clinical Psychology Review*, 20(3), 379–403.
- Shahim, S. (2004). Reliability of the social skills rating system for preschool children in Iran. *Psychological Reports*, 95, 1264–1266.
- Takegami, M., Suzukamo, Y., Wakita, T., Noguchi, H., Chin, K., Kadotani, H., Fukuara, S. (2009). Developmental of a Japanese version of the Epworth sleepiness scale (JESS) based on item response theory. *Sleep Medicine*, 10, 556–565.
- Warnes, E. D., Sheridan, S. M., Geske, J. & Warnes, W. A. (2005). A contextual

Social Skills Scale

approach to the assessment of the social skills identifying meaningful behaviors for social competence. *Psychology in Schools*, 42(2), 173–188.

Wentzel, K. R. (1991). Relations between social competence and academic achievement in early adolescence. *Child Development*, 62, 1066–1078.

Westhues, A., Hanbidge, A. S., Gebotys, R. & Hammond, A. (2009). Comparing the effectiveness of school-based and community-based delivery of an emotional regulation skills program for children. *School Social Work Journal*, 347(1), 74–92.

Woodward, L. J. & Fergusson, D. M. (2000). Childhood peer relationship problems and later risks of educational under-achievement and unemployment. *J. Child Psychiat.* 41(2), 191–201.

Zvolensky, M. J., Strong, D., Vujanovic, A. A. & Marshall, E. C. (2009). Evaluation of anxiety sensitivity among daily smokers using item response theory analysis. *Journal of Anxiety Disorders*, 23, 230– 239.

Yukay-Yuksel, M. (2009). A Turkish version of the school social behavior scale(SSBS). *Educational Science Theory & Practice*, 9 (3), 1633-1650.