

Research Article

Psychometric Properties and Validation of the Arabic Academic Performance Rating Scale

M. F. A. Q. Quadri,^{1,2} M. S. El-Kordy,^{3,4} M. E. Moukhyer,⁵ A. Mukhayer,⁶ and B. Ahmad²

¹Dental Public Health, Department of Quality Assurance, College of Dentistry, Jazan University, Jazan, Saudi Arabia

²Department of Dental Public Health, Universiti Sains Malaysia (USM), Penang, Malaysia

³National Accreditation Authority for Translators and Interpreters, Sydney, NSW, Australia

⁴Department of Quality Assurance, College of Medicine, Jazan University, Jazan, Saudi Arabia

⁵Department of Quality Assurance, College of Applied Medical Sciences, Jazan University, Jazan, Saudi Arabia

⁶Department of Health promotion, Maastricht University, Maastricht, Netherlands

Correspondence should be addressed to M. F. A. Q. Quadri; dr.faeq.quadri@gmail.com

Received 15 February 2017; Revised 14 May 2017; Accepted 4 July 2017; Published 31 July 2017

Academic Editor: Jose C. Nunez

Copyright © 2017 M. F. A. Q. Quadri et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aim. To validate the Arabic version of Academic Performance Rating Scale. **Method.** Translation and test-retest reliability were computed. Exploratory factor analysis and Rasch analysis were conducted to report on the validity. EFA factor structures were evaluated using a Scree plot and the standard multiple criteria included eigenvalue greater than 1. Average measures and step measures were ordered and the mean-square outfit statistic for each category was also evaluated. **Results.** Cronbach's Alpha value of 0.90 was obtained. No differences across category of educational levels were seen ($P > 0.05$). Pattern matrix analysis revealed items to be significantly correlated with each other with a chi square goodness of fit value 0.02. Results supported the interpretation of adequate fit of items, since the infit ZSTD (1.55–0.46), outfit ZSTD (1.59–0.99), infit MNSQ (0.74–1.47), and outfit MNSQ (0.50–1.37) for the items were within acceptable ranges. **Conclusion.** Valid and reliable A-APRS-15 can be used in future studies to see its relation with other variables such as general health and oral health.

1. Background

Education is an integral part of a country's skill based as well as financial based development [1]. The core of teaching and learning experience among the students' seeking education is its assessment. It ascertains their achievement in each subject or course in a curriculum [2]. Standard test score (STS) is one of the popular methods among most institutions for assessment or judgment of a students' academic performance. In Saudi Arabia, assessment is mostly done through standard tests which follow a predesigned curriculum [3]. In this, all the test takers are required to answer similar set of questions which are derived from a common question bank [3–5]. So, standard tests help in attaining scores in a consistent manner through a standardized process. This type of objective assessment is intended to give a fair achievement analysis of each student in their respective courses or subjects. But, it

fails to evaluate the short term and long term needs of a child in a particular course or subject; that is, it does not report on educational experience in a broader perspective [5–7]. It further fails to identify children with learning difficulties so that the nature of support and assistance that they need can be ascertained. It also fails to assist in designing and implementation of appropriate strategies in concordance with specific difficulties that the students may have encountered [6, 8]. Most importantly, some researchers have also raised concern over its reliability [4].

Thus, traditional standard test scores (STS) are solely not enough, and a subjective analysis should be added to complete the overall assessment [9]. Through this, teachers will be able to observe and report on a more distinguished sample of academic content in comparison to traditional standard tests. This will also assist in providing input on social validity data of each student [10, 11]. Few teacher rating

scales were successfully tested and validated in different domains of school experience. Among these are children's behavior rating scale [12], classroom adjustment rating scale [13], teacher child rating scale [14], social skills rating system [15], and behavior and emotional rating scale [16]. These scales are psychometrically sound but possess limitations in monitoring and reporting of academic skill deficits. In addition, the calculation of academic work completion and accuracy rates across each course or subject is also not possible. To overcome this an Academic Performance Rating Scale (APRS) was developed and validated successfully by a group of researchers [17, 18].

APRS (original English version) reflects teachers' perception of a child's academic performance. Initially, thirty items were generated based on suggestions provided by several classroom teachers, school psychologists, and clinical child psychologists. From the original thirty, 19 items were retained with a feedback obtained from another set of teachers, principals, and school and child psychologists regarding item content validity, clarity, and importance [17]. The APRS final version comprised items focusing on one global construct, that is, academic performance through a series of questions (items), like work performance (e.g., estimate the written math work completed relative to classmates) of each child in different subjects; academic success (e.g., what is quality of the child's reading skills?); academic behavior (e.g., how often does the child begin written work prior to understanding the directions?); and academic attention (e.g., how often the child is able to pay attention without you prompting) [17].

Higher educational institutions and graduate producing universities in the Middle East are gradually shifting to English as their medium of instruction due to execution of new educational policies. But, owing to the belief of maintaining and transferring culture and tradition to the coming generations, most national schools (primary and secondary) still follow Arabic as their medium of instruction. On searching the published literature, it is observed that currently there is no subjective assessment scale in Arabic language to assess academic experience of school going children. A questionnaire based assessment if made available will not only reveal the academic/learning experience of Arab children in a broader perspective, but also help in acquiring large-scale analysis to provide quantitative as well as qualitative evidence of the school academic performance in the region. Thus, with this intention, the current study is aimed at providing a validated version of the Academic Performance Rating Scale in Arabic language.

2. Methodology

2.1. Study Setting. This study was conducted in Jazan city of Saudi Arabia, which is situated at southern tip of Arabian Peninsula bordering Yemen. Infrastructure of Jazan is still in development phase unlike Riyadh and Jeddah which are considered as the major cities of Saudi Arabia. Mode of instruction of schools selected in this study is Arabic.

2.2. Arabic Translation. The translation process involved a qualified bilingual translator who carried out the initial

translation of the questionnaire into Arabic language (Arabic-APRS-19, A-APRS-19) while maintaining the conceptual understanding of each question. A second bilingual translator then carried out back translation of the Arabic version into English. A cross examination of the Arabic to English translation found no discrepancies.

2.3. Content Validation. Validation was done to know if the questions are fully assessing the intended outcome (single global construct). A rational analysis of the questionnaire was done by the experts, specifically focusing on the readability, clarity, comprehensiveness, and level of agreement using Likert scale. Three school principals were requested to rate the A-APRS-19 based on the importance of each item (4-point scale: 0 = not at all; 1 = low; 2 = somewhat; 3 = high) and also the consistency between the items (4-point scale: 0 = not at all; 1 = weak; 2 = adequate; 3 = strong) [19].

2.4. Ethical Consideration and Questionnaire Administration. A convenient sample of 100 school children was identified from four different schools in Jazan city of Saudi Arabia. Permission from the review board at Jazan University was followed by permission from school authorities before recruiting the study participants. A face-to-face interview along with a written consent was obtained from both the school teachers and at least one parent of the child. Only healthy children of Arab nationality were included to maintain the homogeneity of the sample. Classroom teachers and children were categorized into 7th grade, 8th grade, and 9th grade, respectively. None of the teachers or child parent refused to be part of the study.

2.5. Statistical Analysis. In order to assess the reliability of A-APRS, internal consistency and test-retest reliability were computed. Internal consistency was measured using Cronbach's α coefficient. Arabic version of APRS (A-APRS) was stated to be internally consistent if it acquired an α coefficient of at least 0.70 [20]. For investigating the stability of A-APRS across times, a test-retest reliability analysis and the intraclass correlation coefficients (ICC) using Pearson's R were calculated (ICC agreements: <0.40, poor to fair; 0.41–0.60, moderate; 0.61–0.80, good; >0.80, excellent) [21]. The distribution of the A-APRS across three different educational levels (7th grade, 8th grade, and 9th grade) was tested to explore discriminant validity and confirm differences, through a nonparametric test (Kruskal-Wallis).

Exploratory factor analysis (EFA) using principal component analysis and Rasch analysis were conducted to report on validity of the questionnaire. As the applicability of Rasch model is dependent on the assumption of unidimensionality (i.e., items defining a single global construct: academic performance), an EFA was first conducted to identify latent constructs (dimensions) in the A-APRS [22]. Later, EFA factor structures were evaluated using a Scree plot and the standard multiple criteria include eigenvalue greater than 1.0 [23]. Average measures and step measures (+SE) were ordered and the mean-square outfit statistic for each category was also evaluated [24]. All the data entry and statistical analysis were done using SPSS version 22 (IBM, USA).

TABLE 1: Background characteristics of study population.

Variables	A-APRS		P value
	Low	High	
Gender	Percentages		
Male	53.84	46.15	0.61
Female	52.63	47.36	
Location			
Urban	55	45	0.53
Rural	50	50	
Mothers' education			
Uneducated	66.66	33.33	0.80
Secondary school	50	50	
High school	53.84	46.15	
Graduate	50	50	
Postgraduate	0	100	
Fathers' education			
Uneducated	0	0	0.23
Secondary school	0	0	
High school	69.23	30.76	
Graduate	46.66	53.33	
Post graduate	25	75	
Mother employed			
Yes	54.54	45.45	0.60
No	52.38	47.61	
Father employed			
Yes	53.12	46.87	NS
No	0	0	
Standard test scores			
Less than 50%	83.33	16.66	0.01*
More than 50%	14.28	85.71	

*Significant with chi square value less than 0.05; NS: not significant.

3. Results

The total sample size was 100, out of which the male school children were 40.6% and the female children were 59.4%. Mean age of study participants was 9.59 ± 1.38 years. In terms of location, 62.5% resided in the urban suburbs, whereas 37.5% belonged to rural suburbs. Previous semesters' standard test scores (STS) were also obtained in order to check its relation with the A-APRS. It revealed that 56.3% of school children were below average (below the optimum requirement). A cross tabulation of A-APRS with other measured variables (parents' education, location, etc.) was performed and the results are reported in Table 1. The students who scored less in the traditional STS have also scored less in their subjective analysis ($P < 0.05$).

3.1. Reliability. Nineteen-item Arabic version of APRS (A-APRS-19) was subjected to reliability tests by assessing internal consistency through Cronbach's Alpha. The value obtained was 0.68 (Table 3) which was less than the expected value of 0.70. On examining the item correlation matrix (matrix of Pearson-type correlations), it was seen that four

of the nineteen items were not in concordance with rest of the items. Linear influence of these variables is corrected by taking them out of the matrix. After removing and recalculating Cronbach's Alpha, a value of 0.90 (Table 3) was obtained indicating that the fifteen-item A-APRS-15 was internally consistent. Descriptive statistics of both A-APRS-19 and A-APRS-15 are displayed in Table 2.

To measure the intraclass correlation (ICC), test-retest reliability was performed. A-APRS-15 was distributed twice for the study sample in two consecutive weeks. Care was taken while repeating the measurement as the original scale (APRS) indicates that the rating for each student is represented only for the previous week [17]. Correlation value of 0.91 showed that there was an excellent agreement between the repeated administrations [25].

3.2. Validity. Currently there is no other Arabic academic rating scale in order to see the relation between them in measuring the same outcome. Administration of A-APRS-19 and A-APRS-15 to the same population revealed a significant correlation. Also, no significant differences in A-APRS-15 across the category of educational levels were seen ($P > 0.05$). A-APRS-15 scored consistently with the obtained standard test scores ($P = 0.01$) (Table 1). Content validation through experts revealed no major corrections based on a mean rating cut-off of 2.5 or higher [19]. Category function statistics revealed a positive analysis of the structural validity of the A-APRS five-point Likert scale; that is, function of assessment was acceptable as each category count was equal to or more than 10.

Model's unidimensionality was confirmed by examining the explained and unexplained variance. Exploratory factor analysis (EFA) revealed that A-APRS-15 across its items measured the outcome (single global construct, that is, academic performance) successfully, thus representing one-factor analysis. Through pattern matrix analysis, it was seen that the items were significantly correlated with each other and a chi square goodness of fit value of 0.02 was obtained. Sampling adequacy was assessed with Kaiser-Meyer-Olkin measure which revealed a value of 0.78 indicating the sample used was nearly adequate. Bartlett's test of sphericity significance value 0.00 indicated equal variances (homogeneity) across the sample.

A single-factor solution was reported using the criterion of eigenvalue larger than 1 (Figure 1). Examining the "percent of variance explained" suggested that this one-factor solution explained 84% of variance. The Scree plot illustrated a pronounced drop after the first factor, further confirming a strong one-factor solution (Figure 1). Unidimensionality and redundancy of the items in the questionnaire were evaluated in a Rasch analysis using the Partial Credit Model as displayed in a tabulated format (Table 4). For the sake of improvement, the misfitting items as well as items minimizing overlap in the level of difficulty represented in the scale were omitted. They were classified as redundant if the infit/outfit mean square (MNSQ) was less than 0.50 or if it was more than 1.37. Results supported the interpretation of adequate fit of items, since the infit ZSTD (1.55–0.46), outfit ZSTD (1.59–0.99), infit MNSQ

TABLE 2: Descriptive statistics of A-APRS (19) and A-APRS (15).

Items	§ A-APRS (19)			§ A-APRS (15)		
	Mean	Std. deviation	*R	Mean	Std. deviation	*R
APRS1	2.25	1.524	.977	2.25	1.524	.858
APRS2	2.56	1.605	.934	2.56	1.605	.907
APRS3	2.59	1.720	.892	2.59	1.720	.847
APRS4	2.72	1.550	.940	2.72	1.550	.743
APRS5	2.75	1.437	.930	2.75	1.437	.906
APRS6	3.34	1.181	.975	3.34	1.181	.875
APRS7	4.19	.965	.948	4.19	.965	.806
APRS8	3.41	.979	.905	3.41	.979	.797
APRS14	3.66	1.234	.925	3.66	1.234	.778
APRS16	2.16	1.081	.931	2.16	1.081	.694
APRS17	3.34	1.335	.870	3.34	1.335	-.729
APRS18	2.69	1.674	.862	DEL	DEL	DEL
APRS19	2.16	1.019	.909	DEL	DEL	DEL
APRS13	1.91	.893	.785	DEL	DEL	DEL
APRS15	2.78	1.338	.914	DEL	DEL	DEL
APRS9	3.53	1.270	.854	3.53	1.270	.797
APRS10	3.47	1.319	.905	3.47	1.319	.851
APRS11	3.91	.995	.933	3.91	.995	.785
APRS12	2.59	1.292	.937	2.59	1.292	-.587

* R: Pearson's coefficient of correlation; § A-APRS (19) and A-APRS (15) intend to measure a single global construct, academic performance; DEL: deleted.

TABLE 3: Scale statistics and reliability of A-APRS (19) and A-APRS (15).

Scale	Mean	Variance	Std. deviation	Cronbach's Alpha
A-APRS (19)	56.00	93.55	9.67	0.68
A-APRS (15)	46.47	164.77	12.83	0.90*

* Cronbach's Alpha value of 0.70 or more was considered significant.

(0.74–1.47), and outfit MNSQ (0.50–1.37) for the items were within acceptable ranges (Table 4).

4. Discussion

This is the first study that attempted to provide a reliable and valid subjective Academic Performance Rating Scale in Arabic. Overall result of both the exploratory factor analysis and Rasch analysis of the data indicated A-APRS-15 (Appendix – B in Supplementary Material available online at <https://doi.org/10.1155/2017/1504701>) as a highly reliable single construct scale. Exploratory factor analysis also provided the preliminary evidence of high internal reliability as evaluated by high alpha statistic. Person and item reliability statistics revealed and confirmed the highly reliable single-factor structure of A-APRS-15.

Initial analysis of A-APRS-19 revealed that items 13, 15, 18, and 19 were not in alignment with the rest of the items. The overall outcome measure was not reliable as Cronbach's Alpha value was less than the expected 0.70. Thus, it was necessary that these items be removed in order to

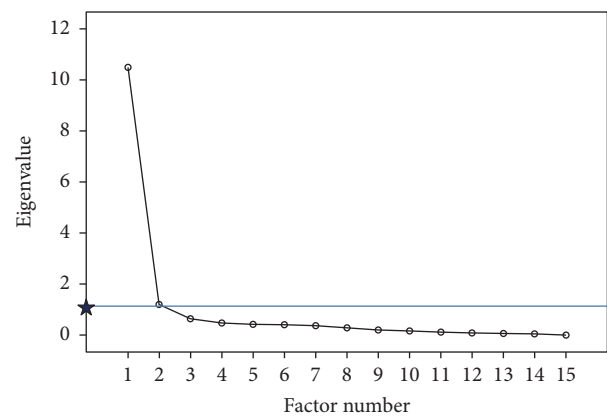


FIGURE 1: Scree plot created through factor analysis. Eigenvalue of 1.0 was considered as the cut-off value for obtaining the Scree plot.

achieve a more reliable scale so that all items contribute towards measuring one outcome (academic performance). Test-retest of A-APRS-15 also revealed a high correlation value which means that A-APRS-15 yielded consistent results over repeated administrations. Rasch analysis was then run on the modified scale termed as A-APRS-15 indicating this to be a validated and reliable single construct scale.

A feedback from the classroom teachers through a verbal survey suggested that most teachers found A-APRS-15 to be a potentially useful tool to measure overall academic performance of a child. The average time required for each teacher to complete the A-APRS-15 was around 8–10 minutes per student. Some teachers also suggested that, apart from

TABLE 4: Rasch analysis of A-APRS-15.

^a Item	Item mean	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
APRS1	0.36	0.78	-1.4	0.50	-0.2
APRS2	0.42	0.85	-1.0	1.37	0.6
APRS3	0.49	1.37	2.30	1.52	0.80
APRS4	0.45	0.85	-1.0	0.82	0.0
APRS5	0.49	1.18	1.0	1.14	0.4
APRS6	0.63	0.74	-1.60	0.54	-1.0
APRS7	0.66	0.71	-1.8	0.77	-0.3
APRS8	0.65	0.83	-0.9	1.36	0.8
APRS14	0.62	0.86	-0.7	0.61	-0.8
APRS16	0.60	0.92	-0.3	0.71	-0.5
APRS17	0.71	0.64	-2.1	0.47	-1.3
APRS9	0.81	0.96	0.0	0.55	-0.7
APRS10	0.82	1.09	0.4	1.20	0.5
APRS11	0.80	1.13	0.5	0.62	-0.3
APRS12	0.89	0.68	-1.2	0.36	-0.6

^aSingle global construct; Partial Credit Model was used for Rasch analysis.

calculating academic experience of major subjects like math, science, and literature, there should be inclusion of other important subjects/courses when developing and validating new academic scales.

Considering the limitations of the current subjective analysis of academic performance, the authors admit that there could be a possible inherent bias. As data was not normally distributed there are possibilities of restricted range in the correlation statistics. The current study does not enforce the subjective assessment as a sole measure for a child's academic experience. On the contrary, it is to inform that this subjective assessment will help in better understanding of areas where the child faces difficulties during his or her learning experience in major subjects so that proper measures are taken, thus lending a helping hand towards a brighter future. A-APRS-15 measure of academic performance will be an easy assessment that can be used in future studies to see its relation with other variables such as general health and oral health.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] N. Alromi, *Vocational Education in Saudi Arabia*, Penn State University, State College, PA, USA, 2000.
- [2] D. Wiliam, "Assessment: the bridge between teaching and learning," *National Council of Teachers of English, Voices from the Middle*, vol. 21, no. 2, article 15, 2013.
- [3] I. A. Al-Sadan, "Educational assessment in Saudi Arabian schools," *Assessment in Education*, vol. 7, no. 1, pp. 143–155, 2010.
- [4] A. Alsadaawi, "Saudi national assessment of educational progress (SNAEP)," *International Journal of Education Policy and Leadership*, vol. 5, no. 11, pp. 1–14.
- [5] W. Popham, "Why standardized tests don't measure educational quality?" *ASCD Educational Leadership Using Standards and Assessment*, vol. 56, no. 6, pp. 8–15, 1999.
- [6] A. Kohn, "Standardized testing and its victims," *Education Week*, 2000.
- [7] J. Verman, P. Aschbacher, and L. Winters, "A practical guide to alternative assessment," ERIC Document Reproduction Service No. ED352389, 2008.
- [8] N. C. o. E. i. Education, *A Nation at Risk: The Imperative for Educational Reform* (1999). Retrieved February 2, 2007, <http://www.ed.gov/pubs/NatAtRisk/index.html>.
- [9] M. Taras, "Assessment: summative and formative-some theoretical reflections," *British Journal of Educational Studies*, vol. 53, no. 4, pp. 466–478, 2005.
- [10] A. Kohn, "Standardized testing: separating wheat children from chaff children," in *What Happened to Recess and Why Are Our Children Struggling in Kindergarten*, S. Ohanian, Ed., McGraw-Hill, New York, NY, USA, 2002.
- [11] P. Seedhouse, "Classroom interaction: possibilities and impossibilities," *ELT Journal*, vol. 50, no. 1, pp. 16–24, 1996.
- [12] M. B. Bronson, B. D. Goodson, J. I. Layzer, and J. M. Love, *Child behavior rating scale*, Abt Associates, Cambridge, MA, USA, 1990.
- [13] R. Barkley, "Child behavior rating scale and check lists," in *Assessment and Diagnosis in Child Psychopathology*, pp. 113–155, Guildford, Newyork, NY, USA, 1988.
- [14] A. W. Hightower, W. C. Work, E. L. Cowen, and C. A. Rohrbeck, "The teacher-child rating scale: a brief objective measure of elementary children's school problem behaviors and competencies," *School Psychology Review*, vol. 15, no. 3, pp. 393–409, 1986.
- [15] F. E. Greisham, "Social skills rating system," *Journal of Psychoeducational Assessment*, vol. 17, pp. 392–397, 1999.
- [16] M. H. Epstein, G. Ryser, and N. Pearson, "Standardization of the behavioral and emotional rating scale: factor structure, reliability, and criterion validity," *The Journal of Behavioral Health Services & Research*, vol. 29, no. 2, pp. 208–216, 2002.

- [17] G. a. R. Du Paul, "Academic performance rating scale," *School Psychology Review*, vol. 20, no. 2, pp. 284–300, 1991.
- [18] B. J. Zimmerman and A. Kitsantas, "Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement," *Contemporary Educational Psychology*, vol. 39, no. 2, pp. 145–155, 2014.
- [19] A. Rice, L. Wakefield, K. Patterson et al., "Development and validation of a questionnaire to measure serious and common quality of life issues for patients experiencing small bowel obstructions," *Healthcare*, vol. 2, pp. 139–149, 2014.
- [20] J. B. Nunnally and I. R. Bernstein, *Psychometric Theory*, McGraw-Hill, New York, NY, USA, 3rd edition, 1994.
- [21] J. J. Bartko, "The intraclass correlation coefficient as a measure of reliability," *Psychological Reports*, vol. 19, no. 1, pp. 3–11, 1966.
- [22] K. E. Green and C. G. Frantom, "Survey development and validation with the Rasch model," in *Proceedings of the International Conference on Questionnaire Development, Evaluation, and Testing*, Charleston, SC, USA, 2002.
- [23] A. B. Costello and J. W. Osborne, "Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis," *Practical Assessment, Research and Evaluation*, vol. 10, no. 7, pp. 1–9, 2005.
- [24] J. M. Linacre, "Optimizing rating scale category effectiveness," in *Introduction to Rasch measurement*, E. V. Smith Jr. and R. M. Smith, Eds., pp. 258–278, JAM, Maple Grove, MN, USA, 2004.
- [25] J. D. Evans, *Straightforward Statistics for the Behavioral Sciences*, Brooks-Cole Publishing, Pacific Grove, CA, USA, 1996.

