



## Linguistic Validation of the Urdu Translation of Slosson Intelligence Test SIT-4 (Vocabulary) Sub-test for Adults: An Educational Perspective

Uzma Qayyum Khan<sup>1</sup>, Muhammad Faran, Ph.D<sup>2</sup>, Noshi Iram Zaman, Ph.D<sup>3</sup>

### ARTICLE INFO

#### Article History:

Received: 6 April 2022

Received in revised form: 7 June 2022

Accepted: 5 October 2022

DOI: 10.14689/ejer.2022.100.004

#### Keywords

Translation, Validation, Slosson Intelligence Test, Vocabulary, University Students.

### ABSTRACT

**Purpose Purpose:** The current aim of the study was to translate and linguistically validate the vocabulary (sub-test) of Slosson Intelligence Test-4 (SIT-4) into the Urdu language for Pakistani adults. **Methodology:** In Phase I, the translation of the vocabulary (sub-test) of SIT-4 was conducted by using standardized forward as well as backward translation, including tryout. In phase II of the study, confirmatory factor analysis was carried out for validation of vocabulary (sub-test) of SIT - 4 using Structural Equation Modeling (SEM). For the psychometric estimation, a sample of 900 (450 males and 450 females) university students aged 19-37 (M =24.38, SD= 3.45), were selected.

**Findings:** The findings of Confirmatory Factor Analysis of the vocabulary (subtest) of SIT-4 resulted in high validity and reliability coefficients. The results showed that the vocabulary (sub-test) of SIT - 4 explained 52% and 53% of the variance for both males and females, respectively. It is concluded that the Urdu version of the vocabulary (sub-test) has high psychometric properties and is culturally and linguistically appropriate for the Pakistani university student population. **Implications for Research and Practice:** It is evident that this study would pave way for future research in intelligence testing, future research can utilize these findings with a diverse age range.

© 2022 Ani Publishing Ltd. All rights reserved.

<sup>1</sup> Corresponding Author, Department of Professional Psychology, Bahria University, Islamabad, Pakistan.  
Email: [uzmaqkhan@gmail.com](mailto:uzmaqkhan@gmail.com), [Orcid: 0000-0003-0750-9910](https://orcid.org/0000-0003-0750-9910)

<sup>2</sup> Department of Professional Psychology, Bahria University, Islamabad, Pakistan.  
email: [mfaran.buic@bahria.edu.pk](mailto:mfaran.buic@bahria.edu.pk), [Orcid: 0000-0003-1200-8846](https://orcid.org/0000-0003-1200-8846)

<sup>3</sup> Department of Professional Psychology, Bahria University, Islamabad, Pakistan.  
email: [noshi.zaman@bahria.edu.pk](mailto:noshi.zaman@bahria.edu.pk), [Orcid: 0000-0002-6556-8030](https://orcid.org/0000-0002-6556-8030)

## Introduction

Since there is evidence from several fields that culture has a significant impact on a person's ability to think critically or their level of intelligence, intelligence is seen as a culturally dependent construct (Panicker, Hirisave, & Subbakrishna, 2006). The demands of the activity itself, as well as the knowledge and traits the individual brings to the task, all have an impact on how well an individual performs on any task. There are several educational and therapeutic uses for intelligence testing, making it crucial to have a thorough, well-standardized intelligence assessment method. To forecast one's learning potential, intellectual capacity must be evaluated. It is also very helpful in the clinical and research domains (Ambreen & Kamal, 2014).

Assessments are now routinely modified to make cross-cultural comparisons since we live in multicultural and multilingual cultures. It is essential to adhere to best practices while modifying tests in order to avoid mistakes and guarantee that test results can be compared. Additionally, it is critical to confirm that the modified test is an accurate and reliable rendition of the original. The effectiveness of the translation and adaptation procedure is essential to guaranteeing the reliability and value of the modified test. The goal of adaptation, which is far from a simple undertaking, is to preserve the content and cultural meaning equivalency between the original and the translated/adapted test. This encourages the comparability of results across persons from various cultural groups (Hernández et al., 2020).

An attempt to linguistically validate the vocabulary subtest of the Slosson Intelligence Test- 4<sup>th</sup> edition (SIT-4) (Slosson, 2017) in the Urdu (native) language of Pakistan was a necessary step. In Pakistan, interest in intelligence testing has increased due to its potential for research, education, and scientific advancement. For hiring and training purposes, several institutions and organizations in Pakistan use intelligence tests. Despite the significance of intelligence testing in Pakistan, scarce studies have been done on test formulation, and there have been no notable initiatives. There have been some noteworthy attempts in the past, but they have not been able to provide worthwhile identification due to methodological and content weaknesses, making them unsatisfactory (Hussain, 2001), and they have not been able to be considered psychometrically appropriate due to insufficient compliance to the standard of test adaptation guidelines (Aziz & Ahmad, 1993).

SIT-4 is a screening test, developed to provide a valid, reliable, and quick estimate of (approximately 10-20 minutes) general verbal cognitive ability which can be administered by individuals not trained in giving individualized tests of general intelligence, thus making its use widespread; with its clinical brevity, psychometric accuracy, ease to administer, and inclusive interpretation structure, it's one of the unique screening test of crystallized intelligence, especially in countries where English language is used. However, its use is restricted in non-English speaking countries due to the language barrier and culturally charged content. It has six cognitive domains that help in screening the crystallized intelligence of children and adults, those categories are vocabulary (VO) having 33 items, general information (GI) has 29 items, similarities and difference (SD) has 30 items, comprehension domain (CO) has 33 items, quantitative (QN) has 34 items, and auditory memory (AM) has 28 items. All of six domains are densely loaded with verbal content which implies the existence of culturally and linguistically inappropriate items.

Also, SIT-4 measures crystallized intelligence which is culturally specific (Weiss et al., 2006). Among all six domains of Slosson intelligence Test- 4<sup>th</sup> edition (SIT-4), vocabulary subtest was chosen as it seems to be the most culturally loaded index.

Instead of creating a brand-new test from scratch when a certain trait or competency has to be measured, researchers frequently translated and modified tests that have been created and validated in other cultures (Hambleton & Lee, 2013). In addition to saving time and other resources for academics, adaptation also makes cross-cultural comparisons easier. Given that people are today immersed in multicultural and multilingual situations, this is an essential aspect (Duarte & Rossier, 2008). Hence, the rationale of the current study is formulation of the Urdu version of the Vocabulary subtest of SIT-4 as no such intelligence screening tool is available in native language, making a dire need for screening tool of intelligence in Pakistan. The Urdu translation was carried out in Phase I of this research and its linguistic validation was conducted in Phase II.

## Method

### Research Design

The present study was conducted in two phases using the correlation (cross-sectional) design. The first phase (phase I) was the translation of Vocabulary (sub-test) of SIT-4 which was carried out by using a standardized forward-backward translation procedure (WHO, 2020). In phase II the linguistic validation of Vocabulary (sub-test) of SIT-4 was conducted. The author's permission for the translation and validation was sought.

### Phase I - Phase I Translation of Vocabulary (sub-test) of SIT- 4.

Translation was carried out by using following standardized forward/backward translation method.

#### Step 1- Forward translation:

For the translation of vocabulary (sub-test) of SIT-4; five bilingual experts were assigned for forward translation. Three experts were PhD scholars in Psychology and two were Master's in Psychology. To ensure the authenticity of forward translations; it was taken into consideration that the members are familiar with the test/item construction as well as knowledge of cultural context. All translators were requested to retain the meaning and item-difficulty of the translation as close as possible to the original test, and easily comprehensible.

*Step 2 - Committee approaches:* After translations, to assess conceptual and connotational equality, comprehensibility, and cultural linguistics, a separate five membered committee was formed. Inconsistencies, sentence formation or syntax, any missing words, ambiguities, or discrepancies were identified and modified by the committee members. In one item from vocabulary domain (Item 149: What does Facsimile mean?), the word "facsimile" was replaced with a fax machine (فیکس مشین) to make it more understandable for the examinee as the word "facsimile" was not commonly used in Pakistani culture. Eventually, a single version was finalized by choosing the best-formatted sentences. In the end, this version of forward translation was matched with the English source version for any mistakes, omitted words, ambiguities, or inconsistencies.

*Step 3- Backward Translation:* After the formation of the final forwarded translation (Urdu), three bilingual experts translated it back into the original language (English). The objective of reverse translation is to identify contradictions and inconsistencies resulting from contextual differences by comparing the original version with the translated material. The three bilingual experts were not the same who had carried out the forward translation and were unaware of the original scale. All the translators were PhD. Scholars in Psychology. After the three backward translations (Urdu-English) were completed, the researchers and the supervisor reviewed and compared them to detect similarities and any incongruities. After finalizing a single backward translation, it was compared with the original test, no substantial discrepancies were found.

*Step 4 - Review and Scrutiny:* After completion and necessary modification of the final form of the Urdu version of the vocabulary (sub-test) of SIT-4 the proofreading and grammatical analysis were carried out. In this final step, all grammatical and language proofreading was done lucratively. No significant incongruities were identified in this step.

*Step 5 - Tryout:* The main objective of this step was to establish the comprehensibility of the Urdu translated vocabulary sub-test of SIT-4. 50 male and female participants were recruited with an age ranging from 19 to 37 years. They were briefed about the aim and nature of the research. The test was administered in a standardized manner, followed by interviews to inquire about the cultural relevance and level of difficulty of the items, and clarity of the instructions. The scoring was done according to the Technical and Interpretative Manual of SIT-4 (2017). The reliability of the subtest was checked through Kuder-Richardson (KR-20) which was .75.

### ***Phase II: Validation of Vocabulary (sub-test) of SIT-4***

This step involved validating the Urdu version of the vocabulary sub-test of SIT-4 and determining its reliability as well as validity (Lamers et al., 2011). This validation took place across both male and female gender. The sample size for this study was based on the recognized fact that to validate a scale better the sample size should be large as well. Hence, 900 participants aged 19 and above were purposely drawn from different institutions of Islamabad and Rawalpindi, Pakistan.

During this step, the factor-structure of the Urdu translated form of the vocabulary (sub-test) of SIT -4 was determined through the psychometric properties, including reliability and validity (convergent).

#### **• *Research Sample.***

A sample of 900 students, including 450 males and 450 females with the age range of 19-37 years ( $M = 22.83$ ,  $SD = 3.11$ ), were recruited for the empirical evaluation of the vocabulary (sub-test) of SIT- 4. The inclusion criteria included both males and females with age ranging from 19 and above, currently enrolled in an educational institute. Exclusion criteria included individuals with visual impairment and learning disabilities, exceptional and intellectually handicap adults (Slosson, 2017). The sample size was determined using the criteria of 10:1 (at least 10 cases per item) proposed by Tabachnick and Fidell (2013). The sample was conveniently drawn from the different institutes of Islamabad and Rawalpindi, Pakistan.

• **Research Instrument and Procedure**

This test was developed by Richard L. Slosson (Slosson, 1963). The vocabulary sub-test reflects the “ability to orally use, understand and define words” as communication skills relies on vocabulary ability. Vocabulary tests, both pictorial and oral, have been a mainstay for rapidly estimating intelligence. There are many tests that measure vocabulary only and most general tests of intelligence include vocabulary subtest. SIT -4 has total 33 items. The scoring of this test is (1) for passing and (0) for failing, the answers given in the manual are objective, and easy to score, the scoring is based on ten in a row correct responses (basal score) and ten in a row incorrect response (ceiling score). The testing ends when the ceiling score is reached. The total raw score is obtained by counting all the passed items from basal to ceiling, giving credit to all the items above the basal score (Slosson, 2017).

• **Data Analysis**

Descriptive statistics were calculated to summarize the demographic characteristics of the participants. Whereas structural equation model (SEM) was carried out to conduct confirmatory factor analysis to determine the psychometric properties of Urdu version of vocabulary (sub-test) of SIT-4 for adults

**Results**

Table 1 shows the descriptive statistics of the sample characteristics. Overall, the sample comprised of 47.33% male and 52.66% female students with the average age of 22.83 years with SD of 3.1 years. Almost 46.11% students were from public schools while 53.88% of the students were from public sectors institutes. Meanwhile 71.88% were undergraduate students while 28.11% were postgraduate students.

**Table 1**

*Descriptive Statistics of the Sample Characteristics (N = 900)*

<b>Variables</b>	<b>f(%)</b>	<b>M(SD)</b>
Age (years)	-	22.83 (3.11)
Gender		
Male	426(47.33)	
Female	474(52.66)	
Institute		
Public	415(46.11)	
Private	485(53.88)	
Education		
Undergraduate	647(71.88)	
Postgraduate	253(28.11)	

Confirmatory factor analysis (CFA) was performed using structural equation modelling (SEM). 33 vocabulary questions from the SIT-4 subtest were subjected to the CFA in order to confirm the factor structure of the translated (Urdu) version. Table 2 lists the model fit metrics for the evaluated model.

**Table 2**

*Confirmatory Factor Analysis, Fit Indices for Vocabulary (Sub-test) of SIT-4 for Adults (N = 900).*

Model	$\chi^2$	df	$\chi^2/df$	GFI	CFI	NFI	RMSEA	SRMR
Initial Model	3214.17	990	3.24	.91	.89	.90	.08	.09
Model Fit	1771.21	986	1.80	.97	.94	.93	.06	.05
$\Delta \chi^2$	1327.85*							

Note. GFI= Goodness of fit index, CFI=comparative fit index, NNFI = non-normed fit index; RMSEA=root mean square error of approximation, SRMR=Standardized root means square,  $\Delta \chi^2$  = chi-square change.

Table 1 displays the fit indices of the vocabulary (sub-test) for both absolute and relative model fit. The first model's absolute fit index revealed that the estimations of the fit were subpar, reading as  $\chi^2 (986) = 1771.21, p < .05$ . In a typical model, the sample size and the number of estimated parameters are thought to have a significant impact on the chi-square statistic, which is used to measure the absolute model fit (Hair et al., 2010). Therefore, in this perspective, researchers advised taking into account various relative fit indices, such as the Goodness of Fit Index (GFI), Cumulative Fit Index (CFI), Normative Fit Index (NFI), Root Mean Square Approximation Error (RMSEA), and Standardized Root Mean Square (SRMR).

Some guidelines were suggested to be followed in order to assess the model's fit; for instance, the  $\chi^2/df$  should vary between 0 and 3. To be deemed excellent estimates for the model, the RMSEA and SRMR estimates must be .08 or less, while the CFI, NFI, and GFI estimates must be .90 or higher (Hu & Bentler, 1999). As a result, the specified criteria for model fit were not met by the present estimations of the relative fit (initial model).

Consequently, the model modification procedure was started in order to achieve the model fit. Therefore, only those covariances between the error terms that had contextual meaning were extracted from the indicators of the measurement model of the vocabulary (sub-test) (Kenny, 2011). Following the drawing of the covariances between the error components, the absolute and relative fit indices were once more compared. The GFI, CFI, and NNFI values were .97, .94, and .93, respectively, while the RMSEA and SRMR were .06 and .05, respectively. As a result, the model fit indices and criteria fell into the category of excellent model fit.

After achieving the stringent criteria of model fit, the factor structure of the vocabulary (sub-test) of SIT- 4 was psychometrically evaluated and reliability and validity (convergent validity) of the vocabulary (sub-test) of SIT- 4 was determined. The investigators suggested that composite reliability and Cronbach's alpha reliability coefficients should be .70 or greater for the constancy of the factor structure while the index of average variance extracted (AVE) should be .50 or greater to claim the good convergence of the measurement (Hair et al., 2010; Henseler, Hubona, & Ray, 2016). The average variance extracted (AVE) is the average of the square root of the factor loading for the respective factor (Hair et al., 2010). The percentage of the variance for the vocabulary (sub-test) of SIT- 4 was 52 and 53 for males and females respectively. However, the reliability coefficients, including composite and Cronbach's alpha, ranged from .91 to .92.

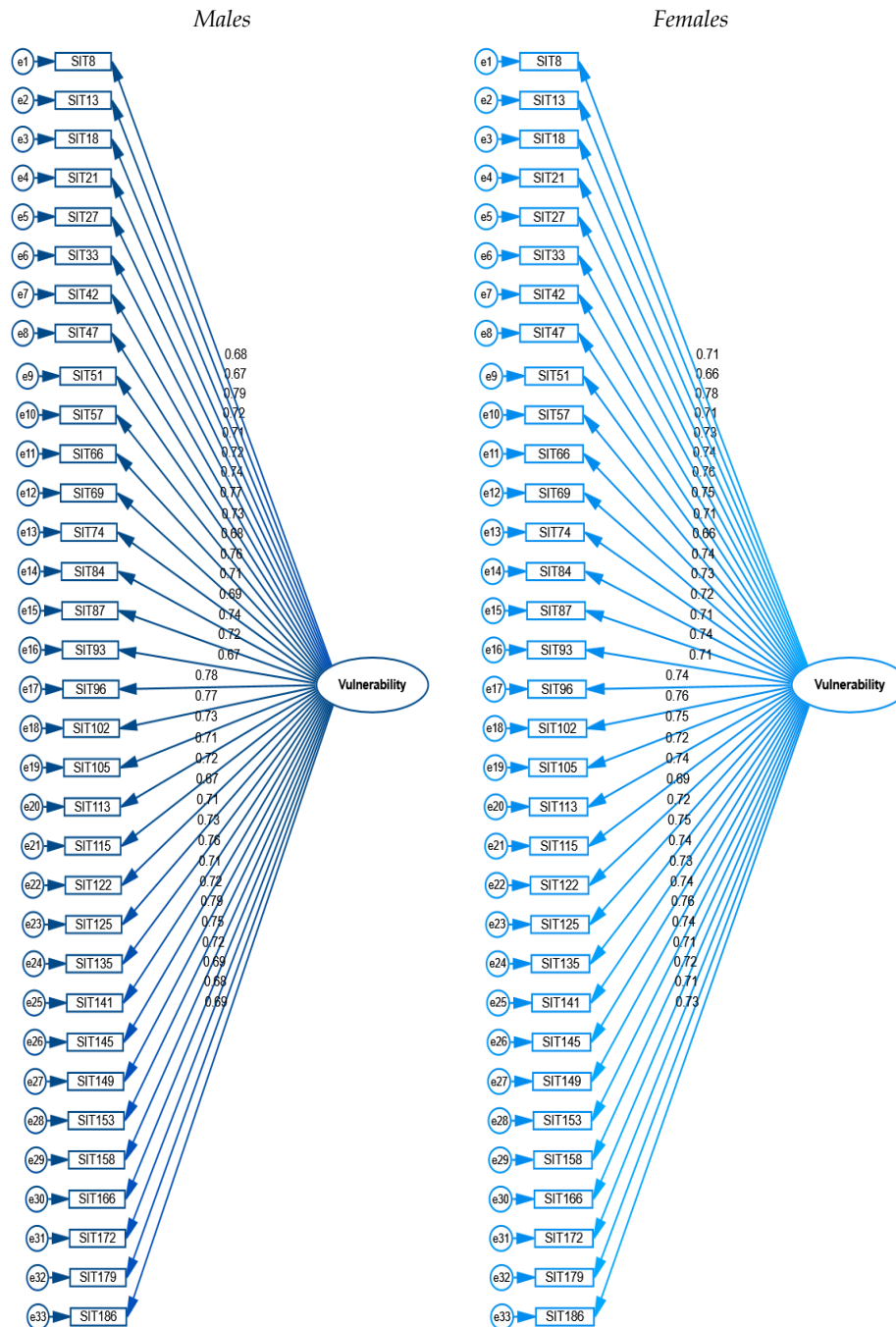


Figure 1: Confirmatory Factor Analysis of Vocabulary (Sub-test) of SIT -4 for Adults (N = 900).



**Table 3**

*Psychometric Evaluation of Vocabulary (Sub-test) of SIT -4 for Adults (N = 900).*

Items	Males			Females		
	a	CR	AVE	a	CR	AVE
SIT8						
SIT13						
SIT18						
SIT21						
SIT27						
SIT33						
SIT42						
SIT47						
SIT51						
SIT57						
SIT66						
SIT69						
SIT74						
SIT84						
SIT87						
SIT93						
SIT96						
SIT102						
SIT105						
SIT113						
SIT115						
SIT122						
SIT125						
SIT135						
SIT141						
SIT145						
SIT149						
SIT153						
SIT158						
SIT166						
SIT172						
SIT179						
SIT186						

Note. CR = Composite reliability, AVE = Average variance extracted,  $\lambda$  (lambda) = standardized factor loading

• **Measurement Invariance for Vocabulary (Sub-test).**

A measurement invariance test was also conducted to determine whether the measure could be used to both males and females The goal of measurement invariance was to assess



whether individuals from various groups of people understand a measure conceptually or contextually alike or not (Byrne & Van de Vijver, 2010). Measurement invariance is divided into two categories: full measurement invariance and partial measurement invariance (configural, metric, scalar, factor variance-covariance, and error variance invariance). Only configural, metric, and scalar invariance were taken into consideration by partial invariance (Hair et al., 2010). In order to assess the measurement invariance, hierarchical models must be used in a series of sequential phases, often starting with the selection of an unconstrained configural model that fits the data well without being divided into groups (see table 4).

**Table 4**

*Testing for the Full Vocabulary (Sub-test) of SIT- 4 for Adults.*

Model	$\chi^2$	df	$\Delta\chi^2$	$\Delta df$	CFI	$\Delta CFI$	RMSEA
Unconstrained	1250.32	493	-		.94	-	.063
Configural Invariance	1771.21	986	520.89	493	.946	.006	.059
Metric Invariance	1811.42	1018	40.21	32	.949	.003	.058
Scalar Invariance	1856.37	1051	44.95	33	.952	.004	.056
Factor Covariance Invariance	-	-	-	-	-	-	-
Error Variance Invariance	1902.12	1085	45.75	34	.952	.002	.052

**Note.** CFI=comparative fit index; RMSEA=root mean square error of approximation, SRMR=Standardized root mean square,  $\Delta\chi^2$  = chi square change,  $\Delta df$ = degree of freedom change,  $\Delta CFI$ = comparative fit index change.

Comparative analysis is done on the restricted models' parameters' invariance between various groups. This measurement invariance process involves comparing each succeeding constrained model to the preceding constrained models. Depending on the degree of invariance, the constraints on the layered model are incrementally introduced, and the models are incrementally analyzed. Theorists recommended using the likelihood ratio test to evaluate the comparability of the stacked models (also known as the chi-square test of difference). In the comparison of two nested models, if the chi-square difference test is statistically non-significant ( $p > .05$ ), it indicates that the two models are invariant across various groups (Hair et al., 2010).

On the other hand, the chi-square test is typically thought to be sensitive to the sample size, number of factors, and non-normality of the distribution (Hair et al., 1999). Cumulative fit index (CFI) (delta CFI), according to Cheung and Rensvold (2002), is some robust criteria for assessing invariance. The limited add to the nested model is permitted if the change in CFI is 0.01 or less.

Primarily, the unconstrained model (un-splitted model) was equated with the well-fitted multi-group (constrained) model, the estimates showed that  $\Delta\chi^2 = 520.89$  with  $\Delta df = 493$  at  $p > .05$ ,  $\Delta CFI$  was .006. That depicted that the factor structure of vocabulary (sub-test) SIT-4 was invariant across males and females. Secondly, metric invariance (equal

factor loading) is considered to be the most crucial invariance in process of measurement invariance. As seen in table 3, when the factor loadings of vocabulary (sub-test) SIT- 4 were constrained across males and females, with  $\Delta\chi^2 = 40.21$  with  $\Delta df = 32$ , the CFI differences between the configural model and the constrained model (metric) did not reach 0.01. This piece of data suggested that metric invariance between both sexes had been established.

The third stage involved determining the scalar invariance of the mean and intercept. Measurement invariance is tested for using scalar invariance, which requires that the means and intercepts of the latent component be the same for all groups. However, scalar invariance was demonstrated, with the CFI of both constrained models (metric and scalar) being less than 0.01 with  $\Delta\chi^2 = 44.95$  with  $\Delta df = 33$ , to reveal invariant means intercepts across both genders. Furthermore, the evidence of error variance invariance also suggested that males and females experienced error variance equally, as the  $\Delta CFI$  was .002,  $\Delta\chi^2 = 45.75$  with  $\Delta df = 34$  at  $p > .05$ . The results showed that the vocabulary (sub-test) SIT- 4 was fully measurement invariant for both males and females.

### Discussion

Test adaptations have multiplied dramatically in recent decades and are now a standard practice (Epstein, Santo, & Guillemin, 2015; Hambleton & Zenisky, 2011). Transcultural adaptation is the process of formulating versions of an assessment instrument that is equal to the original tool while also linguistically and culturally modified (Valer et al., 2015). The items from the vocabulary subtest of SIT-4 were translated into simple and comprehensible Urdu, the target language. One of the items from the Vocabulary sub-test was adapted as the meaning was unclear in the cultural context of Pakistan (Item 149: What does Facsimile mean?), the word "facsimile" was replaced with a fax machine (فیکس مشین) to make it comprehensible for the examinee as the word "facsimile" was not commonly used in Pakistani culture. This contextualization and lucidity was the primary concern in Phase I of the study.

In Phase II, the vocabulary (sub-test) of SIT-4 was linguistically validated across males and females. Estimating a test's psychometric qualities is essential in any applied subject. Any test's strength resides in both its validity and reliability. Psychometric evaluation of the vocabulary subtest showed excellent reliability; the reliability coefficients, including composite and Cronbach's alpha, which ranged from .91 to .92, and the convergent validity of the construct i.e., average variance extracted (AVE) showed excellent estimates as well. For model fitting, the Model modification process was instigated; the model fit indices met the category of excellent model fit. Also, a test of measurement invariance was carried out to evaluate the generalizability of the scale across both genders; The evidences signified the establishment of metric invariance across both genders. Lastly, the invariance of mean and intercept (scalar invariance) was determined, and the results revealed the full measurement invariance of the vocabulary (sub-test) SIT- 4 for both male and females.

### Conclusion

The SIT-4 is an individual screening test of intelligence which can be used for both children and adults, requiring no specialized training and takes up to 15-20 minutes to administer and score. The purpose of the current study was to translate/adapt and validation Vocabulary subtest of SIT-4 in Urdu language to make its use applicable in non-

English speaking country like Pakistan where Urdu is the national language. The study findings revealed good validity and reliability of the Urdu translated/adapted items of the vocabulary sub-test of SIT-4 and proved its validity as all the items showed good psychometric estimates for both males and females. The results of the measurement invariance test showed that the measure was invariant across both males and females for configural, metric, scalar, factor covariance, and error variance. The evidence also showed that the Urdu adapted version of Vocabulary sub-test of SIT-4 can be used in future studies in Pakistan. It can be concluded that the adapted version of SIT-4 vocabulary sub-test will help and improve the screening process of intelligence of the Pakistani children and adults.

### Limitations and Recommendations

There were certain limitations on this study. Firstly, the time boundedness and monetary restraint prevented the development of norms and further strengthening the psychometric properties of the adapted SIT-4. The development of norms for the adapted SIT-4 would enhance its psycho-educational utility for the Pakistani population and a more accurate picture of their intelligence level can be presented. Future studies using the adapted vocabulary subtest of SIT-4 should include a wider age range and geographically diverse sample for more precise representation. Furthermore, other demographic characteristics besides those employed in this study, such as age, education, parental education, etc., can be utilized to determine the psychometric properties of the adapted SIT-4.

### References

- Ambreen, S., & Kamal, A. (2014). Development of Norms of the Adapted Verbal Comprehension Index Subtests of WISC-IV for Pakistani Children. *Journal of Behavioural Sciences*, 24(1), 85-97. [http://pu.edu.pk/images/journal/doap/PDF-FILES/Abstract%20No%207\\_V\\_24\\_No\\_1\\_%202014.pdf](http://pu.edu.pk/images/journal/doap/PDF-FILES/Abstract%20No%207_V_24_No_1_%202014.pdf)
- Aziz, S., & Ahmad, I. (1993). Adaptation of columbia mental maturity scale in pakistan. *Pakistan Journal of Psychological Research*, 8(1-2), 31-42. <https://pjpr.scione.com/cms/abstract.php?id=568>
- Byrne, B. M., & Van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107-132. <https://doi.org/10.1080/15305051003637306>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, 9(2), 233-255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Duarte, M. E., & Rossier, J. (2008). Testing and assessment in an international context: Cross-and multi-cultural issues. In *International handbook of career guidance* (pp. 489-510). Springer. [https://doi.org/10.1007/978-1-4020-6230-8\\_24](https://doi.org/10.1007/978-1-4020-6230-8_24)
- Epstein, J., Santo, R. M., & Guillemin, F. (2015). Cross-cultural adaptation of questionnaires: review of concepts and current guidelines. *Journal of Clinical Epidemiology*, 68(4), 435-441. <https://doi.org/pi.2014.11.021>
- Hair, J. D., Anderson, R. E., Tatham, R. L., & Black, W. C. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Prentice Hall.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, C. B. (1999). *Análisis Multivariante*. Pearson Prentice Hall. <https://dialnet.unirioja.es/servlet/libro?codigo=320227>

- Hambleton, R. K., & Lee, M. K. (2013). Methods for translating and adapting tests to increase cross-language validity. *The Oxford handbook of child psychological assessment*, 172-181. <https://doi.org/10.1093/oxfordhb/9780199796304.013.0008>
- Hambleton, R. K., & Zenisky, A. L. (2011). Translating and Adapting Tests for Cross-Cultural Assessments. In D. Matsumoto & F. J. R. van de Vijver (Eds.), *Cross-Cultural Research Methods in Psychology* (pp. 46-74). Cambridge University Press. <https://doi.org/10.1017/CBO9780511779381.004>
- Henseler, J., Hubona, G., & Ray, P. A. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial management & data systems*, 116(1), 2-20. <https://doi.org/10.1108/IMDS-09-2015-0382/full/html>
- Hernández, A., Hidalgo, M. D., Hambleton, R. K., & Gómez Benito, J. (2020). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema*, 32(3), 390-398. <https://doi.org/10.7334/psicothema2019.306>
- Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Hussain, S. S. (2001). *Development, validation, and standardization of a group verbal intelligence test in Urdu for adolescents*. (Unpublished Doctoral Dissertation). National Institute of Psychology, Quaid-i- Azam University, Islamabad, Pakistan.
- Kenny, D. A. (2011). Respecification of latent variable models. *dauidakenny.net*, 12(2012), 109-119. <http://dauidakenny.net/cm/respec.htm>
- Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). *Journal of clinical psychology*, 67(1), 99-110. <https://doi.org/10.1002/jclp.20741>
- Panicker, A. S., Hirisave, U., & Subbakrishna, D. (2006). WISC-III uk: Comparison of Indian and UK norms. *Journal of Indian Association for Child and Adolescent Mental Health*, 2(4), 108-111. <https://doi.org/10.1177/0973134220060403>
- Slosson, R. (1963). *Manual: Slosson Intelligence Test for Children and Adults*. East Aurora, NY: Slosson Educational Publ. <https://worldcat.org/en/title/25446906>
- Slosson, R. L. (2017). *(SIT-4) Slosson Intelligence Test-4th Edition*. Slosson Educational Publication. [http://www.slosson.com/onlinecatalogstore\\_c394797.html](http://www.slosson.com/onlinecatalogstore_c394797.html)
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics* (6th ed.). Boston, MA: Pearson.
- Valer, D. B., Aires, M., Fengler, F. L., & Paskulin, L. M. G. (2015). Adaptation and validation of the Caregiver Burden Inventory for use with caregivers of elderly individuals. *Revista Latino-Americana de Enfermagem*, 23, 130-138. <https://doi.org/10.1590/0104-1169.3357.2534>
- Weiss, L. G., Saklofske, D. H., Schwartz, D. M., Prifitera, A., & Courville, T. (2006). Advanced clinical interpretation of WISC-IV Index Scores. In L. G. Weiss, A. Prifitera, D. H. Saklofske, & J. A. Holdnack (Eds.), *WISC-IV advanced clinical interpretation* (pp. 139-179). MA, California, London: Elsevier Inc. <https://doi.org/10.1016/B978-012088763-7/50005-X>
- WHO. (2020). *Process of translation and adaptation of instruments*. World Health Organization. [http://www.who.int/substance\\_abuse/research\\_tools/translation/en/](http://www.who.int/substance_abuse/research_tools/translation/en/)