Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

# DEVELOPMENT AND VALIDATION OF THE SITUATIONAL JUDGEMENT TEST ON LEADERSHIP (SJT-L)

Shahida Nasreen[1], Rizwan Akram Rana[2]

## Abstract

*The purpose of this study was to develop and validate a leadership test named Situational Judgement Test on Leadership (SJT-L) to assess the level of leadership skill in adolescent girls studying in a semi-government school of Lahore, Pakistan. The study was conducted in four distinct phases. In first phase, two focus group interviews were conducted to record the "critical incidents" from the context of the population under study. These incidents helped the researchers to develop 17 scenarios and their responses. In phase two and three, a systematic and evidence-based approach developed by Yusoff (2019[a, b]) to conduct content validation and face validation respectively was adopted. Through this intensive process, total fifteen scenarios with five responses were retained. In last phase, SJT-L was administered to 152 girl students of grade VI-X at the same school. Exploratory factor analysis was conducted to categorize the items in SJT-L. After the factor loading analysis, all the 15 items were finalized. Four factors were identified including Self-Management Skill (5 items), Social Engagement Skill (5 items), Collaborative Learning Skill (3 items) and Futuristic Thinking Skill (2 items). The Cronbach alpha value for SJT-L was 0.86 for 15 items. For each sub-scale, the Cronbach alpha value ranged from 0.88 to 0.96. The results indicated that SJT-L is valid as well as reliable and can be used to assess leadership skill in adolescent girls.*

**Keywords:** *student leadership, situational judgement test, semi-government, adolescent girls, critical incidents, subject matter experts (SMEs)*

## Introduction

A variety of methods are being used to assess the leadership skills in participants in a particular setting including Likert scales (Vaughan et al., 2020), Forced-Choice (FC) Scale (Feri, 2015), virtual reality scales (Alcañiz et al., 2018), guided discussions, panel interviews, writing assessments (Biggs et al., 2024) and situational judgement tests-SJTs (Guenole et al., 2015). However, most frequently used Likert scales are being questioned now for indicating potential for cultural biases (Arnulf & Larsen, 2020), response biases, especially in children and adults, putting measurement accuracy at stake (Li et al., 2024) and being vulnerable to reference effects (Marsh & Hau 2003). On the other hand, situational judgement tests (SJTs) have evolved as valuable tools for assessing professional behavior and non-academic skills including social and emotional skills (Murano et al., 2021; Walton et al., 2022), interpersonal relations (Christianet al., 2010) leadership integrity (Rafique & Ghazal, 2022) and leadership (Lievens & Motowidlo, 2016).

Situational Judgment Tests (SJTs) and Likert scales differ in their validation processes and outcomes. SJTs using best-worst response formats may be less related to cognitive ability and personality than initially hypothesized (Rasmussen, 2010). SJTs using likert scales have demonstrated strong relationship with personality measures (Rasmussen, 2010; Whetzel & McDaniel et al., 2016). Expert validation of SJTs with both ranking and rating methods has been

[1] PhD (Edu) Scholar & Corresponding Author, Department of Education, University of Management & Technology, Lahore (Pakistan). f2021095010@umt.edu.pk
[2] Chairperson & Professor, Department of Education, University of Management & Technology, Lahore (Pakistan). rizwan.akram@umt.edu.pk

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

found challenging, having advantages as well as disadvantages (Schubert et al., 2008). The defining of professional behavior and selection of experts remain difficult in SJT validation (Schubert et al., 2008). This process can be improved by dropping items with mid-range means without compromising on validity (Wend et al., 2018). Overall, the validation process for SJTs is more complex and nuanced as compared to traditional Likert scales.

**Objective of the Study**

The current study was conducted to develop and validate a student leadership test named Situational Judgement Test on Leadership (SJT-L) for adolescent girls studying in a semi-government school of Lahore.

**Literature Review**

Situational judgment tests (SJTs) are assessment tools used to examine test takers' reactions to hypothetical scenarios (Patterson & Driver, 2018). They present scenarios related to their daily experiences and ask test takers to select or rank possible responses (Metcalfe & Dev, 2018). SJTs are being used in various researches since the mid-1990s almost. They are designed to measure non-academic attributes and have shown lower adverse impact compared to other selection methods (Patterson et al., 2016). SJTs are low-fidelity simulations (Motowidlo et al., 1990). The logic is explained by Smith et al. (2023) in these words:

> Test specifications (purpose, target audience, attributes of interest) for an SJT are usually determined by authors of the SJT and are specific to their institutional culture or expectations. Therefore, an SJT developed by one institution may not demonstrate acceptable evidence of validity and reliability for use at another institution. (Pg.7)

They are multidimensional in nature as they can measure multiple constructs simultaneously (Chan & Schmitt, 2017; Lievens et al., 2008). For this reason, coefficient alpha is regarded inappropriate for SJTs (Lievens et al., 2008), however, the SJTs review research by Campion et al. (2014) indicates that coefficient alpha was used by 88.4 % studies on SJTs.

A review of SJTs' development studies revealed that mostly such tests consist of 10 to15 scenarios (Husbands et al., 2015; Reinerman-Jones & Teo, 2016) as they require more time to read the scenarios and analyze all the given responses. SJTs have shown promise in leadership assessment and development. Research indicates that SJTs can measure multiple leadership dimensions, including task-focused, relationship-focused, transformational, and developmental leadership (Murase et al., 2020). Traditionally used for personnel selection, SJTs are increasingly being explored for leadership development purposes and SJT scores can be constructed to provide dimension-level feedback useful for leadership development, despite concerns about their reliability (Guenole et al., 2015, 2022). Meta-analyses have established that SJTs assessing leadership and relational skills demonstrate relatively high validity for overall job performance

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

(Christian et al., 2010). The validity procedures for Likert scales and situational judgment tests (SJTs) differ in several ways. Campion et al. (2014) affirm that "number of items" refer to the "number of situations" included in an SJT. Validating an SJT involves several critical steps. In this regard, first step is to develop the "critical incidents" relating to the constructs under study. The relevant literature recommends to establish content through focus group discussions and expert reviews to ensure the test items reflect the intended construct (Nisa et al., 2022; Rafique & Ghazal, 2022). For instance, Nisa et al. (2022) utilized focus groups to identify characteristics of continuous learning, while Rafique and Ghazal (2022) conducted interviews with managers and subordinates to generate context-specific scenarios. The study conducted by Rubio et al. (2003) provided a comprehensive guide on how to conduct a content validity, including calculating relevant indices and objectively assess the content validity of a scale. In another study, Campion et al. (2014) utilized best practices for developing situational judgment tests, including writing scenarios and response options. Yusoff (2019[a]) has devised a systematic approach to affirm content validity. This approach follows six steps of content validation starting with the development of content validation form, appointment of a panel of experts, content evaluation by panel experts, review of domain and items, scoring each item as per criteria and calculating item content validity index (I-CVI). A higher I-CVI means high consensus between/ among experts on the relevance of the items to the constructs under study ensuring content validity of the instrument. A lower I-CVI indicates that the item may not be much relevant to the construct under measurement. Similarly, another work of the same researcher, Yusoff (2019[b]), has also outlined a well-defined and structured process to achieve face validity of an instrument. The literature suggests to use real applicants or incumbents as respondents (Camion et al., 2014). This process follows six steps wrapping up with the calculation of I-FVI (item face validity index). The participants of High I-FVI scores indicate that a majority of the respondents or raters agree that the item appears clear, appropriate, and understandable. Lower scores suggest the item might need revision for clarity or comprehensibility from the respondents' perspective.

Most of the self-assessed leadership instruments developed for assessing leadership skills in young population are based on Likert type scale. Some widely used Likert type instruments include YLLSDS (Youth Leadership Life Skills Scale) conceptualized by Miller (1976) and developed by Seevers et al. (1995) for American youth, LSI (Leadership Skills Inventory for Students) developed by Karnes and Chauvin (1985), RRSL (Roets' Rating Scale for Leadership) developed by Roets (1986) for grade 5 to grade 12 Chinese students, LPI (Leadership Practices Inventory) developed by Kouzes and Posner (1988) in America for specific use with college students, LSS (Leadership Skills Scale) by Ogurlu and Emir for 6$^{th}$ to 8$^{th}$ grade Turkish students, YLPS (Youth Leadership Potential Scale) developed by Yuan et al. (2019) for 7 to 9 grade Chinese students and MFL-Q (Multifactor Leadership Qualities and Competencies Scale) designed by Manaware (2023) for pre-university and undergraduate students in Bangalore. In the context of resilience, which is related to leadership, the Child and Youth Resilience Measure and the Connor-Davidson Resilience Scale were identified as the most adequate for adolescent populations (Ballard et al., 2023). In a recent study conducted in Pakistan by Imran et al. (2023), a scale was developed based on the Medical Leadership Competency Framework to gather

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

Pakistani medical students' perceptions of leadership training. Only two studies (Abrahim, 2011; Ali, 2011) were conducted in Pakistan at school level to explore the perceptions of student leaders, non-student leaders, principal and the teachers about student leadership development in private secondary schools of Karachi, Pakistan. Both studies were qualitative in nature.

Research studies on leadership development in developing nations emphasize the significance of considering cultural contexts when evaluating leadership practices. Conventional Western-centric frameworks are insufficient for comprehending leadership in diverse cultural settings (Tian, 2022). Cross-cultural research, like the GLOBE project, has uncovered notable variations in leadership across nations, underscoring the necessity for a flexible and collaborative approach to leadership studies (Jepson, 2009). Quantitative evaluations of leadership development levels, taking into account cross-cultural factors, have demonstrated differences among countries, with the highest levels found in Finland and the USA, and the lowest levels in Vietnam and North Korea (Blyznyuk, 2022). These findings underscore the necessity of incorporating cultural context when studying and developing leadership in developing countries.

Pakistani youth at the school level shows a diverse range of leadership qualities influenced by various factors, including gender, cultural backgrounds, socioeconomic status, and educational opportunities. This diversity can manifest in different leadership styles reflecting the unique experiences of students from urban and rural settings. In this regard, SJTs have been labelled as more beneficial than Likert scale as they are aligned with cultural and contextual settings of the participants (Weng et al., 2018). They have high face validity and show low chances of getting fake response (Kasten et al., 2018; Lievens et al. 2008). They have low vulnerability to reference effects demonstrating strong predictive and face validity in educational contexts (Lievens & Sackett, 2012). Despite having some inadequacies like cognitively challenging and having lower reliabilities (Lievens et al. 2008), SJTs are valid in assessing social and behavioral constructs (Smith et al., 2023). Therefore, developing a comprehensive, understandable and applicable SJT to examine the leadership skill in adolescents would be highly beneficial to foster research on student leadership in a developing country's context.

## Methodology

The study used a comprehensive approach to develop and validate SJT-L. A conceptual understanding of this approach is given in Table 1.

In Phase I of the study, as suggested by Campion et al. (2014), the situation construction of SJT-L and response construction was initiated. The process involved two focus group interviews; one with a group of experienced teachers teaching in a semi-government girls' high school, the research site, and other with the girl students studying in the same school in Grade VI to X. In first focus group, there were four teachers and one extra & co-curricular Incharge of the same school. All the teachers had atleast 10 years' experience of teaching at the same school. The objective of these focus group interviews was to record the "critical incidents" in which adolescent girls studying in Grade VI to X were able to develop their leadership skills in the school set-up.

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

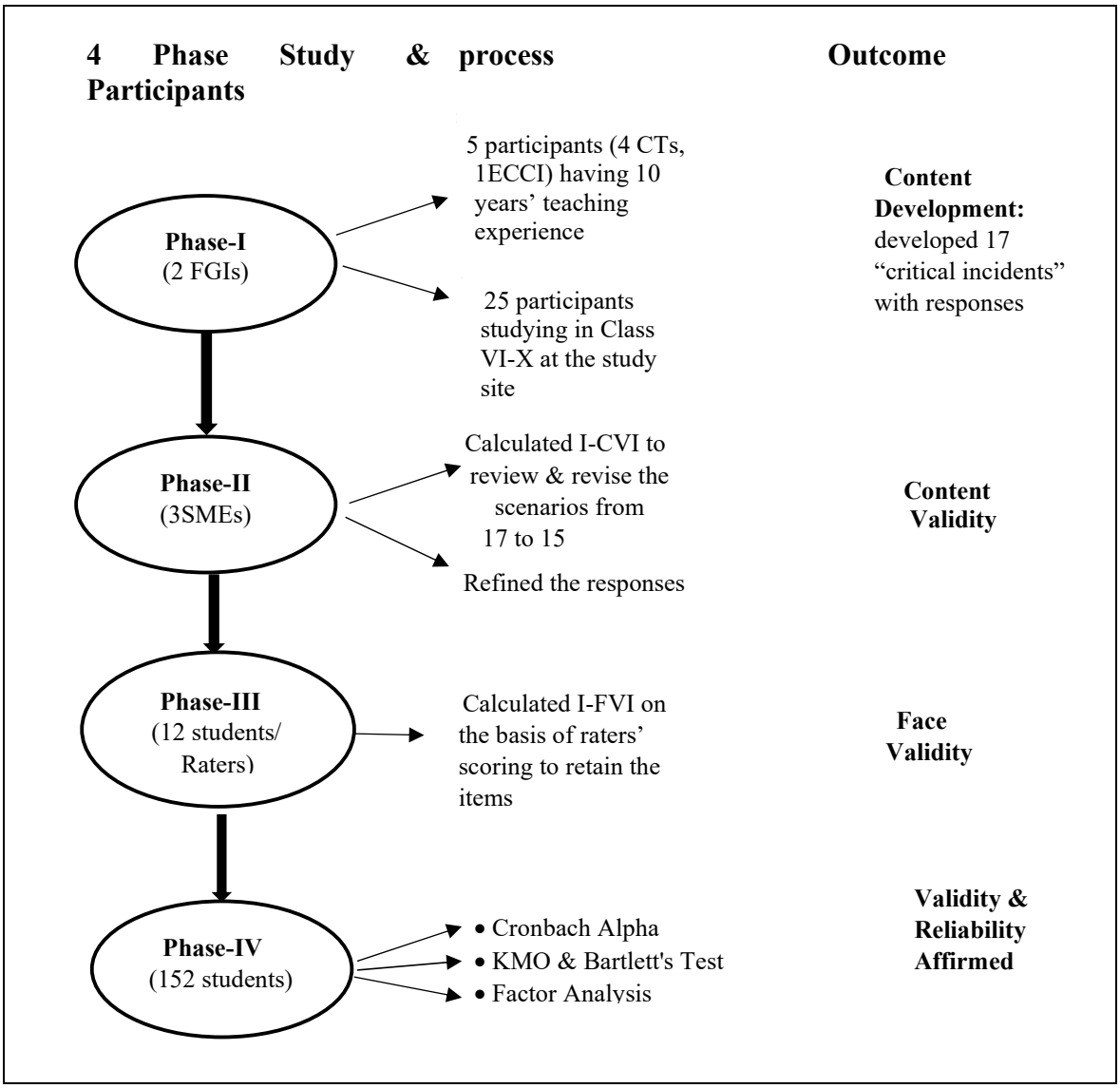
These incidents helped the researchers to develop 17 possible scenarios and their responses related to the constructs of student leadership. As suggested by Souza et al. (2021), the responses were graded under the instructions provided by performance level rubrics which indicated five performance levels for each skill.

The extent to which the components of a scale reflect the specific attributes being measured is called its validity (Cook & Beckman, 2006; Haynes et al., 1995). Pursuing the objective of the study, a systematic and evidence-based approach developed by Yusoff (2019[a]) to conduct content validation by computing content validity index was adopted.

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

**Figure 1**

*Conceptual Framework of the Study*




In Phase II of the study, content validity of SJT-L was established involving three subject matter experts (SMEs); one leadership specialists, one assessment specialist and one expert in developmental psychology. All the experts had atleast ten years' experience in their area of specialization. The SMEs were asked to rate the scale from 1 to 4 (1 for not relevant to 4 for highly

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

relevant to the measured domain). On the basis of the scores awarded by the SMEs, Table 1 was developed.

**Table 1**
*The Relevance Ratings of Items of SJT-L by Three Experts*

| Item No. | Expert 1 | Expert 2 | Expert 3 | Experts in Agreement | I-CVI | UA |
|---|---|---|---|---|---|---|
| 1. | 1 | 1 | 1 | 3 | 1 | 1 |
| 2. | 1 | 0 | 0 | 1 | 0.33 | 0 |
| 3. | 1 | 1 | 1 | 3 | 1 | 1 |
| 4. | 1 | 1 | 1 | 3 | 1 | 1 |
| 5. | 1 | 1 | 1 | 3 | 1 | 1 |
| 6. | 1 | 1 | 1 | 3 | 1 | 1 |
| 7. | 1 | 1 | 1 | 3 | 1 | 1 |
| 8. | 1 | 1 | 1 | 3 | 1 | 1 |
| 9. | 1 | 1 | 1 | 3 | 1 | 1 |
| 10. | 1 | 1 | 1 | 3 | 1 | 1 |
| 11. | 1 | 1 | 1 | 3 | 1 | 1 |
| 12. | 1 | 1 | 1 | 3 | 1 | 1 |
| 13. | 0 | 1 | 0 | 1 | 0.33 | 0 |
| 14. | 1 | 1 | 1 | 3 | 1 | 1 |
| 15. | 1 | 1 | 1 | 3 | 1 | 1 |
| 16. | 1 | 1 | 1 | 3 | 1 | 1 |
| 17. | 1 | 1 | 1 | 3 | 1 | 1 |
| | | | | S-CVI/Ave | 0.92 | |
| Proportion Relevance | 0.94 | 0.94 | 0.88 | S-CVI/UA | | 0.8 |
| Average proportion of items judged as relevance across three experts | | | | | 0.92 | |

*Note.* I- CVI = Item-content validity index; S-CVI =Scale-content validity index of items; S-CVI/Ave= Scale-content validity index of items average;
S-CVI/UA= Scale-content validity index of items average of universal agreement (UA)

As mentioned in Table 1, I-CVI values of two items, item 2 and 13, were below acceptance level, so they were removed from the test, SJT-L. By looking at other values including S-CVI/Ave (0.92), S-CVI/UA (0.88) and relevance across the experts (0.92), it is concluded that SJT-S meets

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

acceptance level. After content validity check, total items for SJT-L were limited to 15 scenarios which achieved acceptable content validity.

As the language of SJT-L is English, so it was very essential to find out its face validity for non-native girl speakers of English studying in class VI to X. In order to achieve the objective, a systematic and evidence-based approach developed by Yusoff (2019[b]) to conduct face validation through face validity index was adopted.

**Table 2**

*The Clarity and Comprehension Ratings on the SJT-L by 12 Raters*

| Item No. | Raters in Agreement | I-FVI | UA |
|---|---|---|---|
| 1. | 12 | 1 | 1 |
| 2. | 12 | 1 | 1 |
| 3. | 12 | 1 | 1 |
| 4. | 12 | 1 | 1 |
| 5. | 12 | 1 | 1 |
| 6. | 12 | 1 | 1 |
| 7. | 12 | 1 | 1 |
| 8. | 12 | 1 | 1 |
| 9. | 12 | 1 | 1 |
| 10. | 12 | 1 | 1 |
| 11. | 12 | 1 | 1 |
| 12. | 12 | 1 | 1 |
| 13. | 12 | 1 | 1 |
| 14. | 10 | .83 | 0 |
| 15. | 11 | 0.92 | 0 |
| | S-FVI/Ave | 0.97 | |
| | S-FVI/UA | | 0.87 |
| Proportion of scenarios' clarity & comprehension from rater 1-5, 7-9 &11-12 = 1; rater 6 = 0.86, rater 10=0.93 | | | |
| Average proportion of clarity & comprehension for SJT-L across the 12 raters = 0.99 | | | |

*Note.* I- FVI = Item-face validity index; S-FVI =Scale-face validity index of items; S-FVI/Ave= Scale-face validity index of items average; S-FVI/UA= Scale-face validity index of items average of universal agreement (UA)

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

In Phase III, face validity of SJT-L was determined by employing the test to twelve girl students of class VI-X studying at the research site. The participants were asked to rate the scale from 1 to 4 (1 for not clear, not understandable to 4 as very clear, very understandable). On the basis of the scores awarded by the raters, Table 2 was developed.

Based on the calculations mentioned in Table 2, it can be concluded that S-FVI/Ave (0.97), S-FVI/UA (0.87) and average proportion of clarity & comprehension for SJT-L across the 12 raters is 0.99. It is concluded that SJT-L meets acceptance level, and thus the test comprising of 15 scenarios has achieved face validity. The test is again reviewed and refined (Item 14 & 15 especially) under the suggestions of the raters.

In Phase IV of the study, the refined test was employed to 152 randomly selected girl students of class VI-X studying at the research site. It was a small level pilot study. Data was analyzed in SPSS-21. In order to find out sample adequacy of the data for factor analysis, initially KMO and Bartlett's test was applied, as mentioned in Table 3, given below.

**Table 3**
*KMO and Bartlett's Test*

| KMO and Bartlett's Test | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | .78 |
| Bartlett's Test of sphericity | Approx. Chi-Square | 2391.201 |
| | df | 105 |
| | Sig. | .000 |

*Note.* *($p \leq .05$)

Higher KMO values (closer to 1) indicate adequacy of data for factor analysis. A KMO of 0.78 advocates that the sample is adequate and tends to yield reliable results. A significant result ($p < 0.05$), as you have with $p = 0.000$, indicates that the correlation matrix is not an identity matrix, meaning there are relationships among the variables, and factor analysis can be performed.

Next, the exploratory factor analysis (EFA) technique was applied by using principal components as a method. Varimax rotation method was used. Factor loading values above than 0.5 were accepted as loading values, as mentioned in Table 4. A total of 15 items of SJT-L were confirmed with 4-factor solutions. The results presented in table 4 direct that four factors constituted the constructs of SJT-L.

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

**Table 4**

*Factor Loadings*

| Items | Self-Management Skill | Social Engagement Skill | Collaborative Learning Skill | Futuristic Thinking Skill |
|---|---|---|---|---|
| Responsibility-taking Skill | .777 | | | |
| Time-regulation Skill | .956 | | | |
| Goal-achieving Skill | .857 | | | |
| Decision-making Skill | .898 | | | |
| Anger-control Skill | .941 | | | |
| Conversation-initiating Skill | | .886 | | |
| Argument-presenting Skill | | .868 | | |
| Expression-communicating Skill | | .869 | | |
| Initiative-taking Skill | | .837 | | |
| Active-listening Skill | | .862 | | |
| Empathy-developing Skill | | | .959 | |
| Moral-analysis Skill | | | .939 | |
| Teamwork Skill | | | .948 | |
| Creative-thinking Skill | | | | .918 |
| Abstract-thinking Skill | | | | .927 |

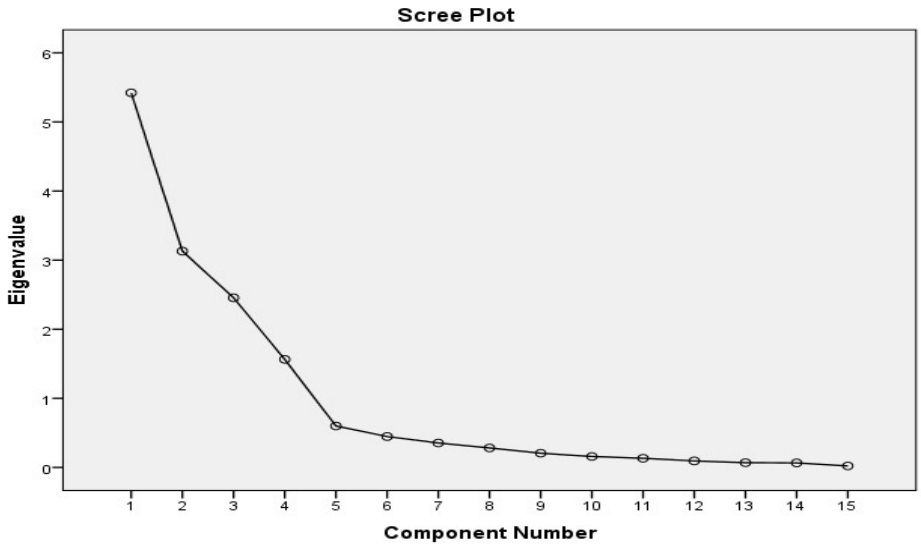
According to table 4, there were finally four factors consisting of total 15 items/ scenarios. The first factor was labelled as "Self-Management Skill" and was loaded with five sub-skills including Responsibility-taking Skill, Time-regulation Skill, Goal-achieving Skill, Decision-making Skill and Anger-control Skill. "Social Engagement Skill" was the second main skill constituted by five sub-skills named as Conversation-initiating Skill, Argument-presenting Skill, Expression-communicating Skill, Initiative-taking Skill and Active-listening Skill. The third main factor was "Collaborative Learning Skill" comprising of Empathy-developing Skill, Moral-analysis Skill and Teamwork Skill. The fourth and last factor was "Futuristic Thinking Skill" constituted by Creative-thinking Skill and Abstract-thinking Skill. The reliability analysis, Cronbach alpha value, for SJT-L was found to be 0.86 for 15 items. Item to total correlation values of all statements were above 0.350.

The scree plot illustrated in Figure 2 presents the Eigenvalues in relation to the items of the scale.

**Figure2**

*Scree Plot presenting the Eigenvalues in relation to the items of SJT-L*

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024




Analyzing the scree plot is instrumental in factor reduction. The examination reveals that the first four Eigenvalues are substantially elevated above the flat line, exceeding the threshold of 1. Beyond this point, the line approaches a horizontal alignment, indicating that four distinct factors are loading prominently.

## Conclusion

SJT-L was developed using "critical incidents" from the life of the participants gathered through focus group interviews. It was demonstrated that the test had acceptable level of both, content validity and face validity. As a result of exploratory factor analysis, four factors were identified. Total items of the scale were 15. The reliability analysis, Cronbach alpha, verified the test is reliable. Item to total correlation values of all statements were above 0.350. The results of the study deduced that the SJT-L had good validity and reliability. Researchers and scholars can assess the level of leadership skill for adolescent girls in semi-government schools of Lahore through this test. The results of the study provide insight to the school administrators, head teachers and teachers, focusing on different sub-skills of student leadership. Further research can be conducted to develop different situational judgement tests to be used in diverse contexts.

## References

Abrahim, A. (2011). Exploring students and teachers beliefs about student leadership and how educational processes foster or hinder the development of student leadership skills in two private secondary schools of Karachi, Pakistan. (MA Thesis at *Aga Khan University, Institute for Educational Development, Karachi.)*

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

Alcañiz, M., Parra, E., Giglioli, I. A. C., & García, A. (2024). Using Virtual Reality for Leadership Assessment and Training through Behavioral Biomarkers. *Biometrics and Neuroscience Research in Business and Management: Advances and Applications*, 141.

Ali, M. (2011) Exploring the beliefs and practices that contribute to youth leadership development in a private higher secondary school in Karachi, Pakistan. (MA Thesis at *Aga Khan University, Institute for Educational Development, Karachi.)*

Arnulf, J. K., & Larsen, K. R. (2020). Culture blind leadership research: how semantically determined survey data may fail to detect cultural differences. *Frontiers in psychology*, *11*, 176.

Biggs, A., Johnston, S., & Russell, D. (2024). Leadership and communication: how to assess executive skills. *Journal of Business Strategy*, *45*(3), 199-205.

doi: 10.1108/JBS-05-2023-0085

Blyznyuk, T. (2022). Assesment of leadership development: cross-cultural aspects.

doi: https://doi.org/10.32782/2524-0072/2022-35-21

Campion, M. C., Ployhart, R. E., & MacKenzie Jr, W. I. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance*, *27*(4), 283-310. https://doi.org/10.1080/08959285.2014.929693

Campos, M. I. D., & Rueda, F. J. M. (2020). Authentic leadership: development and initial validation of a situational judgment test. Revista Psicologia Organizações e Trabalho, 20(2), 1047-1056. https://doi.org/10.17652/rpot/2020.2.18100

Chan, D., & Schmitt, N. (2017). Situational judgment tests. *The Blackwell handbook of personnel selection*, 219-242.

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*(1), 83-117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, *119*(2), 166-e7. doi:10.1016/j.amjmed.2005.10.036

Ferry, K. (2015). Assessment of leadership potential. *Korn Ferry*. http://www.kornferry.com/

Guenole, N., Chernyshenko, O., Stark, S., & Drasgow, F. (2015). Are predictions based on situational judgement tests precise enough for feedback in leadership

e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

development?. *European Journal of Work and Organizational Psychology*, *24*(3), 433-443. doi:10.1080/1359432X.2014.926890

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological assessment*, *7*(3), 238. https://psycnet.apa.org/doi/10.1037/1040-3590.7.3.238

Husbands, A., Rodgerson, M. J., Dowell, J., & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgement test for medical school admissions. *BMC medical education*, *15*, 1-9. doi: 10.1186/s12909-015-0424-0.

Jepson, D. (2009). Studying leadership at cross-country level: A critical analysis. *Leadership*, *5*(1), 61-80. https://doi.org/10.1177/1742715008098310

Karnes, F. A. Chauvin, J. C. (1985). Leadership Skills Inventory. Technical Manual. Scottsdale, AZ, USA: Great Potential Press: DOK Publishers. ISBN 0910707383, 9780910707381

Kasten, N., Freund, P. A., & Staufenbiel, T. (2018). Sweet little lies. *European Journal of Psychological Assessment*. https://psycnet.apa.org/doi/10.1027/1015-5759/a000479

Kouzes, J. M. & Posner, B. Z. (2012). The leadership challenge: How to make extraordinary things happen in organizations (5.Ed). San Francisco: Jossey-Bass. ISSBN 1119278961, 9781119278962

Li, L., Zhang, B., Sun, T., & Drasgow, F. (2024). The More Contextualized, the More Valid: Effects of Contextualization Strategies on Forced-Choice Measurement. *Journal of Business and Psychology*, 1-19. doi: 10.1007/s10869-024-09983-2

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, *9*(1), 3-22. https://psycnet.apa.org/doi/10.1017/iop.2015.71

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, *97*(2), 460. doi: 10.1037/a0025741.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, *37*(4), 426-441. https://doi.org/10.1108/00483480810877598

Marsh, H. W., & Hau, K. (2003). Big-fish-little-pond-effect on academic self-concept. A cross-cultural (26 country) test of the negative effects of academically selective schools. American Psychologist, 58(5), 364-376. doi: 10.1037/0003-066x.58.5.364.

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

Metcalfe, David & Harveer (2020) *Oxford Assess and Progress: Situational Judgement Test*, Third Edition. Oxford Academic. https://doi.org/10.1093/oso/9780198805809.001.0001

Miller, R. A. (1976). Leader/agent's guide: Leadership life skills. Stillwater: Oklahoma State University.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. Journal of Applied Psychology, 75, 640–647. doi: 10.1037/0021- 9010.75.6.640

Murano, D., Lipnevich, A. A., Walton, K. E., Burrus, J., Way, J. D., & Anguiano-Carrasco, C. (2021). Measuring social and emotional skills in elementary students: Development of self-report Likert, situational judgment test, and forced choice items. *Personality and Individual Differences*, *169*, 110012. doi:10.1016/j.paid.2020.110012

Murase, T., Roebuck, A., & Takahashi, K. (2020). Development and validation of a situational judgement test of Japanese leadership knowledge. In *Leadership in East Asia* (pp. 69-92). Routledge. https://doi.org/10.1080/13602381.2018.1548543

Nanaware, R. B. (2024). Youth Leadership Qualities and Competencies: Scale Development and Preliminary Validation. *SPAST Reports*, *1*(1).

Nisa, Y. F., Umar, J., Arista, P. D., Widaningrum, H., & Hayat, B. (2022). Development and Validation of Situational Judgement Test to Measure Continuous Learning Competency. doi:10.15408/jp3i.v11i1.25302

Ogurlu, Ü. (2012). Liderlik becerileri geliştirme programının üstün zekâlı olan ve olmayan öğrencilerin liderlik becerilerine etkisi [Effect of leadership skills development program on the leadership skills of gifted and non-gifted students] (Yayımlanmamış doktora tezi). İstanbul Üniversitesi, İstanbul.

Patterson, F., & Driver, R. (2018). Situational judgement tests (SJTs). *Selection and recruitment in the healthcare professions: Research, theory and practice*, 79-112. doi: 10.1007/978-3-319-94971-0_4

Patterson, F., Knight, A., Dowell, J., Nicholson, S., Cousans, F., & Cleland, J. (2016). How effective are selection methods in medical education? A systematic review. *Medical education*, *50*(1), 36-60. https://doi.org/10.1111/medu.12817

Rafique, A., & Ghazal, S. (2022). Content Validation of a Situational Judgment Test of Manager's Integrity: A Qualitative Approach. *Journal of Professional & Applied Psychology*, *3*(2), 306-318. https://doi.org/10.52053/jpap.v3i2.113

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

Rasmussen, J. L. (2010). *Situational judgment test responding: best and worst or rate each response* (Doctoral dissertation, Texas A & M University).

Reinerman-Jones, L., & Teo, G. (2016, March). Assessing multidimensional complex decision making with situational judgment tests. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 49-55). IEEE. doi: 10.1109/COGSIMA.2016.7497785

Roets, L. S., (1986). Leadership: A skills training program. New Sharon, LA: Leadership Publications.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social work research*, *27*(2), 94-104. https://doi.org/10.1093/swr/27.2.94

Schubert, S., Ortwein, H., Dumitsch, A., Schwantes, U., Wilhelm, O., Kiessling, C. & Kiessling, C. (2008). A situational judgement test of professional behaviour: development and validation. *Medical teacher*, *30*(5), 528-533. doi: 10.1080/01421590801952994.

Seevers, B. S., Dormody, T. J., & Clason, D. L. (1995). Developing a scale to research and evaluate youth leadership life skills development. Journal of Agricultural Education, 36(2), 28-34. doi: 10.5032/JAE.1995.02028

Smith, K. J., Reed, B. N., Neely, S., Farland, M. Z., Haines, S. T., & Robinson, J. D. (2023). Opening the Black Box: Agreement and Reliability of a Situational Judgment Test Across Multiple Institutions. *American journal of pharmaceutical education*, *87*(11), 100129. doi: 10.1016/j.ajpe.2023.100129.

Souza, M. C., Margalho, É. M., Lima, R. M., & Mesquita, D. (2021). Development process of a rubric for assessment of leadership competences in project management scenarios. In *International Symposium on Project Approaches in Engineering Education* (Vol. 11, pp. 151-158). https://doi.org/10.5281/zenodo.509560

Tian, Q. (2022). Strengthening Global Leadership Studies. In *Handbook of Global Leadership and Followership: Integrating the Best Leadership Theory and Practice* (pp. 1-25). Cham: Springer International Publishing. doi: 10.1007/978-3-031-21544-5_56

Vaughan, R., Romanick, M., Brassil, D., Kost, R. G., Schlesinger, S., & Coller, B. S. (2020). 4037 assessing leadership skills in translational science training: The Rockefeller University leadership survey. *Journal of Clinical and Translational Science*, *4*(s1), 116-117. doi: 10.1017/cts.2020.353

Vol. 4, No. 4 (2024)
EDUCATIONAL RESEARCH AND INNOVATION (ERI)
e-ISSN:2710-4354
p-ISSN:2076-9660
Received: 13/06/2024 Accepted: 04/11/2024

Walton, K. E., Burrus, J., Murano, D., Anguiano-Carrasco, C., Way, J., & Roberts, R. D. (2022). A big five-based multimethod social and emotional skills assessment: The mosaic™ by ACT® social emotional learning assessment. *Journal of Intelligence*, *10*(4), 72. doi: 10.3390/jintelligence10040072

Weng, Q. D., Yang, H., Lievens, F., & McDaniel, M. A. (2018). Optimizing the validity of situational judgment tests: The importance of scoring methods. *Journal of Vocational Behavior*, *104*, 199-209. https://psycnet.apa.org/doi/10.1016/j.jvb.2017.11.005

Whetzel, D. L., & McDaniel, M. A. (2016). Are Situational Judgment Tests Better Assessments of Personality than Traditional Personality Tests in High-Stakes Testing?. *The Wiley handbook of personality assessment*, 205-214. https://psycnet.apa.org/doi/10.1002/9781119173489.

Wolcott, M. D., Hahn, F., McLaughlin, J. E., & Cox, W. (2022). Interested in situational judgment tests? Preparing pharmacy educators for potential challenges. *Currents in Pharmacy Teaching and Learning*, *14*(6), 785-789. https://doi.org/10.1016/j.cptl.2022.06.010

Yuan, Y., Chen, Q., Sun, X., Liu, Z., Xue, G., & Yang, D. (2019). Development and preliminary validation of the youth leadership potential scale. *Frontiers in psychology*, *10*, 2310. https://doi.org/10.3389/fpsyg.2019.02310

Yusoff, M. S. B. (2019a). ABC of Content Validation and Content Validity Index Calculation. Education in Medicine Journal, 11(2), 49–54. https://doi.org/10.21315/eimj2019.11.2.6

Yusoff, M. S. B. (2019b). ABC of response process validation and face validity index calculation. *Educ Med J*, *11*(10.21315). doi: 10.21315/eimj2019.11.3.6