Girls' Education Program (GEP)

2023-2026 Impact Evaluation in Nepal: Baseline Report

Fernanda Gándara

April 2024

Acknowledgements

The author wants to thank the following colleagues for their support with the finalization of this report: Sagar Neupane, Sunita Shakya, Mini Joshi, Dhiraj Anand, Reema Shrestha, Avicia Rodgers and Daniel Oliver.

Table of contents

Girls' Education Program: 2023-2026 Impact Evaluation in Nepal Baseline Report

Executive Summary

Room to Read's Girls' Education Program (GEP) is a school-based intervention that seeks to help girls stay in secondary school longer and build the skills to make key life decisions. Within Nepal and across all countries that GEP supports, the program operates through four primary components: life skills education, mentoring, targeted material support, and family and community engagement.

Room to Read Nepal has implemented GEP in Nepal since 2001, with notable evolutions of the program since its initial implementation. The program began as a material support-focused initiative; today, life skills education, centered on themes of collaboration, critical thinking, decision-making, leadership, and resilience, is at the core of the program, along with elements of teacher training, gender responsive teaching, alumnae engagement, psychosocial counseling, and more.

Country-wide, Nepal's students are experiencing a persistent gender gap in educational attainment and learning outcomes, despite overall progress in the educational sector for students. Girls experience intersectional struggles in completing schooling, with completion rates negatively correlated with household poverty, presence of functional difficulties, and whether the girl lived in a rural area. Girls are also experiencing an attainment gap in learning areas that put them at a disadvantage in the current labor market (e.g., information technology). Child marriage, household chores, poverty, and low agency continue to be barriers for girls in school, and girls are reporting discrimination in both school and home contexts. They also report struggles with self-image, socialization, and socio-emotional support.

Given these girl-specific struggles in accessing and excelling at school, as well as making healthy and informed decisions for themselves to lead a fulfilling life, it is not

only critical to implement programs like Room to Read's GEP – these programs need to be evidence-based and effective.

The GEP 2023-2026 Impact Evaluation in Nepal is designed to determine the impact that the program has on the intended outcomes. Specifically, the evaluation aims to determine: 1) The impact of three years of life skills sessions and mentoring on girls' life skills levels; 2) The impact of three years of life skills sessions and mentoring on girls' gender knowledge and attitudes; 3) The impact of the program on girls' advancement levels; and 4) The extent to which girls believe that the different elements and variations of the program are contributing to the intended outcomes.

Data collection for the baseline evaluation was conducted in October 2023, reaching 710 students in schools where Room to Read's GEP is currently being implemented, and 672 students in control schools. Enumerators utilized the updated Adolescent Life Skills Assessment (ALSA) to measure life skills levels of participants. The ALSA focuses five domains: Decision-Making, Emotional Resilience, Leadership, Collaboration, and Gender Knowledge and Attitudes.

Findings from the baseline survey showed significant differences in Collaboration and Emotional Resilience between treatment and control groups. However, upon completing a regression analysis, all significant differences disappeared, fully explained by student- and school-level characteristics (such as school size and urbanicity).

Another notable finding was that there are significant and positive correlations between all the ALSA domains, which is to say, improvements in one domain is generally associated with improvements in the other domains. Further, school size and age were both negatively correlated with performance across domains.

As this evaluation moves into its next stages, the team will take learnings from this baseline to inform design improvements. Psychometric analyses suggest that

increasing the number of items per domain and reconsidering scoring methodologies could improve the reliability of the scores on the ALSA. The team will also continue to review similar projects to improve the design of further stages, such as drawing on projects that suggest differentiating between minor and major decision-making habits.

Girls' Education Program: 2023-2026 Impact Evaluation in Nepal Baseline Report

Introduction

Room to Read's Girls' Education Program (GEP) is a school-based intervention that seeks to improve the educational and life outcomes of women across the globe. GEP's core programming supports girls in secondary schools by providing them with life skills education, mentoring, targeted material support, and by engaging their parents and community members in their educational journey. To date, the program has supported more than 3.4 million girls across eight countries.

In 2023, Room to Read's Research, Monitoring and Evaluation (RME) Department launched a new evaluation of the program in Nepal. The purpose of the evaluation is to gauge the extent to which the program is impacting girls' life skills and secondary advancement rates. This report describes the purpose of the evaluation, its methodology, and the results of the baseline data collection conducted in October 2023 in Nepal.

Bridging Gaps, Facing Challenges: Moving Towards Gender-Inclusive Education in Nepal

In recent years, Nepal has made significant progress in access to education and learning outcomes. The country's primary enrolment rate has increased to 96% and literacy rates of Nepali adults has grown steadily to 68% in 2018 (Babu Adhikari, 2021). However, challenges remain. While access and completion rates have improved, the latter remains strikingly low for upper secondary education (27%) (UNICEF, 2022). Importantly, results are unevenly distributed between groups of students. Across educational levels, completion rates are related to urbanicity and wealth, being much higher for wealthy urban students, especially in upper secondary schooling. Similar

patterns are observed for learning outcomes such as foundational reading and numeracy (UNICEF, 2022).

A particular problem in Nepal is that gender gaps persist in many areas. A higher percentage of girls than boys repeat school years, drop out or do not transition successfully from primary to secondary schooling (UNICEF, 2022). Not surprisingly, these rates intersect with students' characteristics such as wealth, urbanicity and functional difficulties. In addition, girls lag behind boys in learning areas such as Information and Communication Technology Skills (ICT) (UNICEF, 2022), which are critical for accessing current labor opportunities. More importantly, aggregated numbers mask the fact that girls continue to face targeted difficulties in their educational journeys. Demand-side barriers for girls include child marriage (37% by age 18), household chores, poverty, and low agency (Foundation for Development Management, 2021). Supply-side barriers include lack of gender-responsive school environments, problems related to planning, and recurrent governance disruptions (Foundation for Development Management, 2021). Today, adolescent girls in Nepal, even in urban areas, report being discriminated at home and school, as well as high rates of self-hatred (Shreshta, 2023) They report having few friends outside school, limited opportunities for networking and socialization, and little psychosocial support to cope with their challenges (Shreshta, 2023).

Promises and Paradoxes of Life Skills Education in Nepal

There have been several efforts to address the persistent needs of female students in Nepal, including an increased focus on life skills education. The Nepalese secondary curriculum and corresponding teacher preparation include a provision for life skills components. These provisions are generic, such that life skills are learned in a cross-curricular manner (Munsi & Guha, 2014). While Nepal has tried to strengthen the provision of life skills education in secondary schooling through teacher training,

teachers, particularly female teachers, lack the preparation and support to enhance their students' life skills and emotional well-being (Shreshta, 2023).

Recent donor-funded projects supporting marginalized girls in Nepal have shown some promising trends in terms of socio-emotional and life skills outcomes. In the Kailali district, a combination of interventions targeting girls' learning outcomes, life skills (including financial literacy), financial opportunities, school infrastructure, teachers' practices, and networking opportunities yielded significant gains in learning outcomes and transition rates within secondary education (Babu Adhikari, 2021). However, safeguarding mechanisms (e.g., how girls report harassment incidents) were not necessarily enhanced through the interventions. While there were improvements over time, such as more girls reporting incidents to parents, these were not statistically different to the improvements made by girls in control schools. In addition, girls who were part of the program reported higher levels of self-confidence in dealing with adverse situations such as harassment, bullying, abuse, or gender-based violence, partly attributed to self-defense training. Last, the endline evaluation of the project identified some contradictory results in relation to girls' decision-making abilities, with higher levels of decision-making control reported in the quantitative surveys than in the qualitative studies. The results of the evaluation suggest that decision-making tends to be done by parents, particularly fathers (Babu Adhikari, 2021).

Similarly, in Western Nepal[1], an aid project targeting at-risk and out-of-school girls introduced a series of improvements to teaching methodologies, management systems, and school environments. Through a robust package of interventions, which included peer mentoring networks, increased access to role models, life skills education (including financial education), increased access to financial opportunities, girls' clubs, teacher capacity and support, gender sensitive training, and community engagement, among others, the evaluation of the project identified promising results

[1] Districts of Dhading, Lamjung, Parsa, and Surkhet.

for multiple outcomes (Foundation for Development Management, 2021). By endline, a greater number of girls reported having increased life choices, and the levels of child marriage decreased. However, self-initiated marriages – meaning marriages arranged by girls themselves – remained high. The evaluation did not show significant changes in the degree to which girls made or participated in critical decisions in their lives (Foundation for Development Management, 2021). However, statistically significant improvements were made for specific decisions, including the decision to continue studying, the age of choice to get married, whether to work after completing studies and the type of work intended to engage in (Foundation for Development Management, 2021). Last, the project found significant improvements in the time spent on household chores, which decreased to 1.7 hrs. per day by endline – a statistically significant reduction from prior measures of 1.8 hrs. of household work per day.

Room to Read's Girls' Education Program in Nepal

Room to Read's GEP has been implemented in Nepal since 2001, with over 11,000 girls benefitting from this programming as of 2023. In that time, programming has evolved from materials-based support focused on lower grades to a robust approach involving life skills education, family and community engagement, psychosocial counseling, teacher training with a child rights perspective, mentoring, gender responsive teaching, and engagement with alumnae.

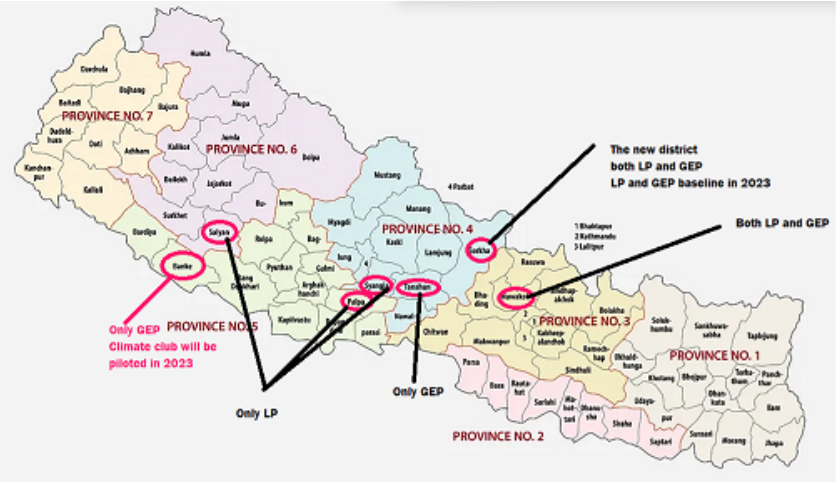
Room to Read's Nepal GEP team works closely with the federal government, as an active member of the Girls and Inclusive Education Network (GIEN) under the Center for Education and Human Resource Development, actively participating in related meetings and providing direct support in developing GIEN's strategy. The GEP life skills manuals have been accredited as national supplementary life skills resource material through Nepal's Curriculum Development Center. Nepal's GEP is currently being implemented in four districts: Banke, Gorkha (new), Nuwakot and Tanahun.

Room to Read participates in Municipal Education Section of Municipalities events and programs.

Though program staff is engaged with the government and ties are strong, the implementation modality is direct, meaning that the entire responsibility of the implementation lies with Room to Read's Country Office as opposed to being shared with Government and other counterparts. Current efforts aim to strengthen local schools and governments' ability to deliver parts of the program, to transition into a more collaborative mode of program implementation.

Figure 1

Room to Read's Implementation in Nepal



In Nepal, Room to Read's Alumnae tend to marry at later ages than national averages and in general, have low levels of disagreements with their families (Gandara, Anand, & Oliver, 2024). Room to Read estimates that less than half of Nepal's GEP Alumnae are married within five years of graduation, and that a large percentage of them are enrolled in tertiary education and/or are working after one year (77.8%) or five years (59.3%). However, disagreements with families over important decisions

persist in some areas such as studying and where to live, to name a few (Gandara, Anand, & Oliver, 2024).

Methodology

The purpose of the GEP 2023-2026 Impact Evaluation in Nepal is to determine the impact that the program has on the intended intermediate outcomes, as per its theory of change. By conducting this evaluation, the team will aim to respond to the following evaluation questions:

1. What is the impact of three years of life skills sessions and mentoring on girls' life skills levels?
2. What is the impact of three years of life skills sessions and mentoring on girls' gender knowledge and attitudes?
3. What is the impact of the program on girls' advancement levels?
4. To what extent do girls believe that the different elements and variations of the program are contributing to the intended outcomes?

To answer the first two questions, the evaluation will follow a repeated cross-sectional design with control groups, wherein the team will examine the changes in outcomes between girls who did not participate in the program (baseline) and girls who have been enrolled in the program for three years (endline) in GEP schools. The team will compare these changes to the changes observed across equivalent cohorts in control schools. Using a difference-in-difference approach, the team will assess the extent to which changes observed in GEP schools are significantly different to changes observed in control schools.

To respond to the third question, the team will use a repeated cross-sectional design with control groups but focusing on a different cohort of students (Cohort 2). Advancement levels will be assessed using the results on Nepal's Secondary Education Examinations at the aggregated level. Last, to respond to the fourth evaluation question, the team will conduct a stand-alone qualitative study of girls' experiences.

The study is expected to take place in 2025 and will be designed in close collaboration with the Nepal GEP team.

Table 1

Evaluation Design – 2023-2026 GEP Evaluation

	2023	2024	2025	2026
G5 (Cohort 1)				
G6				
G7 (Cohort 2)			Girls' Study (Q4)	
G8				Endline (Q1 & Q2)
G9	Baseline (Q1 & Q2)			
G10		Baseline (Q3)		Endline (Q3)

Sample

Initially, the team aimed to reach 800 girls in the treatment and control schools for each data collection point[2]. At baseline, only 710 students in the treatment schools and 672 students in the control schools were successfully reached.

Table 2

Baseline Sample

Sample	Baseline	
	Treatment	Control
Schools	**25**	**40**
Students	**710**	**672**

Based on historical data collected by Room to Read in Nepal and other countries, the team had made the following power assumptions: alpha level of 5%; intra-cluster correlation of 5% (school-level); effect size of 0.25 standard deviations for each

[2] Based on historical parameters and using calculations in Teerenstra et al. (2012).

continuous outcome; cluster autocorrelation of 0.5; person autocorrelation of 0 (cross-sectional design); and average cluster sizes of 25. These parameters suggest that at baseline, the power achieved to compare independent means between treatment and control groups was almost 0.99. By endline, the team will aim to reach the initial target of 800 girls by group, to reach an estimated power close to 0.80 for a difference-in-difference analysis (Teerenstra, Eldridge, Graff, de Hoop, & Borm, 2012)[3].

Revised ALSA

The 2023-2026 GEP Impact Evaluation in Nepal is the first to use the updated Adolescent Life Skills Assessment (ALSA) to measure life skills levels of participants. The updated ALSA is administered using paper and pencil self-assessments, and focuses on five domains: Decision-Making, Emotional Resilience, Leadership, Collaboration and Gender Knowledge and Attitudes. The development of the ALSA items follows a bottom-up participatory design and are designed to reflect girls' experiences and opinions. Extensive development and pilot work predating this evaluation began in November 2022. Field work was conducted in Nuwakot, at that time, and later in Tanahun, in May 2023. A full technical report on the ALSA, its development processes and its technical characteristics will be published later in 2024.

Data Collection

Data collection for the baseline evaluation was conducted in October 2023 in the districts of Tanahun and Gorkha. Data collection was conducted by 9 enumerators (RtR), and data entry was conducted by an external firm whose work was monitored by Room to Read's Nepal Country Office. No major problems were identified.

A sub-sample of 123 treatment girls was reassessed three weeks later. Data from these girls was used to compute test-retest reliability estimates.

[3] Current sample would yield a power of 0.73 for a Difference-in-Differences analysis.

Data Analysis

Data from baseline data collection was cleaned, appended, and analyzed using Stata and R. The cleaning process consisted of removing duplicates and/or empty cases and labeling all variables and values in the dataset. Scoring of the data was conducted on Stata and further psychometric analysis, including reliability estimates and factor analyses were conducted in R.

Human Subjects' Protection

All Room to Read's life skills evaluations go through strict ethics review and are expected to meet global-level and country-level requirements for human subjects' protection. Nepal's evaluation was included in Room to Read's global IRB protocol, with tools, consent forms and assent forms reviewed in September 2023.

Prior to data collection, a consent form was sent to parents of potential participants in both treatment and control schools. Only students with signed consent were invited to participate in the study. In addition, prior to administering the ALSA, students were informed about their rights and only those students who assented to be included in the study took the test.

Results

Result #1. The domains of collaboration and emotional resilience showed significant differences between treatment and control groups. However, these differences disappeared after controlling for student and school variables.

The first result concerns the equivalence of treatment and control groups at baseline. At first glance, there are significant differences between these groups for two domains. As shown in Table 3, means tend to be higher for the control group than the treatment group and the difference was statistically significant for the domains of collaboration and emotional resilience. These differences explained that at the aggregated level, the total score for control girls was significantly higher than that of treatment girls (see Table 3).

Table 3

Treatment and Control Scores by Construct

Domain	Treatment			Control		
	N	Mean	Standard Dev	N	Mean	Standard Dev
Collaboration***	**710**	**6.48**	**1.73**	**672**	**6.72**	**1.64**
Decision-making	**710**	**5.53**	**1.57**	**672**	**5.46**	**1.55**
Emotional Resilience***	**710**	**7.41**	**1.81**	**672**	**7.80**	**1.78**
Leadership	**710**	**7.52**	**1.92**	**672**	**7.67**	**1.75**
Gender Knowledge and Attitudes	**710**	**9.22**	**2.37**	**672**	**9.38**	**2.38**
Total ALSA	**710**	**36.16**	**6.12**	**672**	**37.04**	**5.72**

*** p-value < 0.01

To understand the nature of these differences, a regression analysis was conducted on each of the domains. Each ALSA domain was regressed on the type of school (control vs. treatment), the size of the school[4] (large or not), district (Tanahun or

[4] Large schools were set as those having 31 or more students per classroom.

Gorkha), the interaction between the school and the district, and the age of the girls. The regression analyses revealed that after controlling for these variables, all significant differences disappeared. For example, at the aggregated level, age, district, and the interaction term explained all differences for the total score (see Table 4). Ultimately, younger girls, students in Gorkha, and students in smaller schools tended to have higher ALSA scores.

Table 4

Regression Coefficients – Baseline Evaluation Data

TOTAL ALSA - Baseline	Coefficient	Standard Error	t value	P > |t|
Type (1= Treatment)	-0.60	0.38	-1.57	0.12
Age	**-1.06**	**0.16**	**-6.71**	**0.00**
Large School (1=Yes)	-0.03	0.51	-0.05	0.96
District (1= Gorkha)	**1.49**	**0.48**	**3.11**	**0.00**
District * Large School (1= Large School in Gorkha)	**-2.06**	**0.72**	**-2.88**	**0.00**
Constant	**54.11**	**2.31**	**23.39**	**0.00**

Result #2. ALSA data is multidimensional, yet high correlation is shown between its domains.

Table 5 shows the correlation indexes between the different domains and student- and school-level variables. As observed in Table 5, there are significant and positive correlations between all the ALSA domains. In other words, improvements alongside one of the domains is generally associated with improvements in the other domains. In addition, older ages and girls enrolled in larger schools tended to perform more poorly across all domains. Last, there were more large schools in Tanahun, and students in that district tended to underperform peers in Gorkha, everything else equal.

Table 5

Correlation Between Constructs and Students' Characteristics – All Sample (only significant correlations shown).

Sample	Age	District	Large	Collaboration	Decision-making	Emotional Resilience	Leadership	Gender Knowledge and Attitudes	Total ALSA
Age	**1.00**								
District (1=Gorkha)	**0.11**	**1.00**							
Large School (1=Yes)		**-0.26**	**1.00**						
Collaboration	**-0.06**		**-0.10**	**1.00**					
Decision-making	**-0.09**		**-0.14**	**0.31**	**1.00**				
Emotional Resilience	**-0.18**			**0.20**	**0.19**	**1.00**			
Leadership		**0.09**	**-0.15**	**0.25**	**0.41**	**0.14**	**1.00**		
Gender Knowledge and Attitudes	**-0.14**			**0.27**	**0.28**	**0.33**	**0.18**	**1.00**	
Total ALSA	**-0.17**		**-0.11**	**0.62**	**0.68**	**0.55**	**0.62**	**0.71**	**1.00**

While the different domains are correlated, ALSA data is multidimensional. Here, there are three stronger factors accounting for most of the variance and three smaller factors accounting for most of the remaining variance (see Appendix A.2). Dimensionality considerations, as well as considerations about reliability (see Appendix A.3) and information (see Appendix A.4) will be used to determine whether at endline, scores should be compared at the construct level or at the aggregate level.

Discussion

The 2023-2026 GEP Impact Evaluation in Nepal was designed to evaluate the extent to which Room to Read's GEP is making a difference in participants' life skills levels and related outcomes (e.g., passing rates in gatekeeping exams). Baseline data collection was completed successfully in 2023 and data analysis shows that there are some preexisting differences between control and treatment girls across some of the ALSA domains. These differences are surprisingly consistent with those observed in previous GEP evaluations (e.g., see Anand, 2018). Importantly, these differences are fully explained by student- and school-level characteristics, highlighting the importance of collecting more background data at endline.

In addition, the updated ALSA performed much better than previous versions of the assessment. However, psychometric analyses suggest that there are a few improvements that need to be made by endline to guarantee a reliable and meaningful comparison of scores at the domain level. Adding a few items per domain and reconsidering the scoring methodologies could improve the reliability of the scores and the information provided for each ALSA domain[5].

Finally, the team will continue to review similar projects to improve the design of the upcoming stages of this evaluation. For example, the evaluation of a similar program revealed that decision-making seems to be a layered construct, with minor decisions following a different pattern that big life decisions, where parents seem to have a higher level of influence (Foundation for Development Management, 2021). This is consistent with the results of our own Alumnae Survey, and the team will use this information to explore the different layers of decision-making during the Girls' Study planned for 2025. The Girls' Study will also explore topics such as menstruation taboos and bullying, relevant for Room to Read schools and population (Khan et al.,

[5] See Appendixes for full detail on these psychometric analyses.

2018). Overall, the team is engaged in continuous improvement around measurement and design to maximize our ability to learn and make a difference through the 2023-2026 GEP Impact Evaluation.

References

Adhikari, A. (2021). *Endline Evaluation of Supporting the Education of Marginalized Girls in Kalali (Stem II) Project.* Mercy Corps. https://girlseducationchallenge.org/media/1v4lhdaa/stem-gect-endline-evaluation.pdf

Anand, D. (2019). *Nepal Baseline Life Skills Assessment Report 2019 (Draft Report).* Room to Read.

Bergenfield, I., Jackson, E., & Yount, K. (2019). *A gender-equitable school index for secondary schools in Nepal and beyond.* GAGE. https://www.gage.odi.org/publication/a-gender-equitable-school-index-for-secondary-schools-in-nepal-and-beyond/

Foundation for Development Management. (2021). *Report on Sisters for Sisters' Education-II Project Endline Evaluation.* FCDO - Girls Education Challenge. https://girlseducationchallenge.org/media/nhnlxmym/sisters-for-sisters-gect-endline-evaluation.pdf

Gandara, F., Anand, D., & Oliver, D. (2024). *Alumnae Survey: Summary Tables 2023.* Room to Read.

Khan, Z., Jackson, E., Clark, C., Bergenfeld, I., Puri, M., & Yount, K. (2018). *Supporting girls' education: An evaluation fo Room to Read in Nepal.* GAGE. https://www.gage.odi.org/publication/supporting-girls-education-nepal-es/

Munsi, K., & Guha, D. (2014). Status of Life Skill Education in Teacher Education Curriculum of SAARC Countries: A Comparative Evaluation. *Journal of Education & Social Policy, 1 (1)*, 93-99. https://jespnet.com/journals/Vol_1_No_1_June_2014/13.pdf

Shreshta, B. (2023). *Why emotions matter: Promoting female teachers and gender-transformative social-emotional learning in Nepal.* Center for Universal Education at Brookings. https://www.brookings.edu/articles/why-emotions-matter-promoting-female-teachers-and-gender-transformative-social-emotional-learning-in-nepal/

Teerenstra, S., Eldridge, S., Graff, M., de Hoop, E., & Borm, G. (2012). A simple sample size formula for analysis of covariance in cluster randomized trials. *Statistics in Medicine, 31*, 2169–2178. https://pubmed.ncbi.nlm.nih.gov/22495809/

Thapa, P., Sharma, P., & Pandey, A. (2024). The Landscape of Life Skill Education in Nepal: Current Realities and Future Horizons (A Meta-Synthesis). https://www.researchgate.net/publication/378659426_The_Landscape_of_Life_Skill_Education_in_Nepal_Current_Realities_and_Future_Horizons_A_Meta-Synthesis

UNICEF. (2022). *MICS-Eagle Nepal Education Fact Sheets | 2022.* UNICEF. https://data.unicef.org/wp-content/uploads/2023/06/Nepal_factsheet_Mar_2023.pdf

Annexes

A.1. Distributions of each construct

The Nepal ALSA domains show great distribution and lower levels of skewness than previously collected data on life skills (see Anand, 2018). The highest skewness is observed for Emotional Resilience Scores; the team will consider a few score corrections prior to endline data collection.

Figure 2

Distribution of ALSA Collaboration Scores

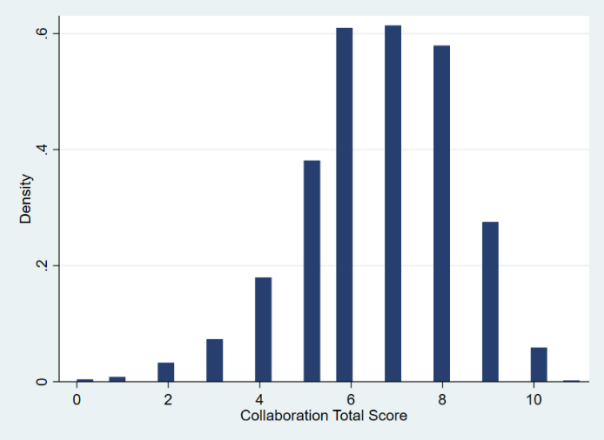


Figure 3

Distribution of ALSA Decision-making Scores

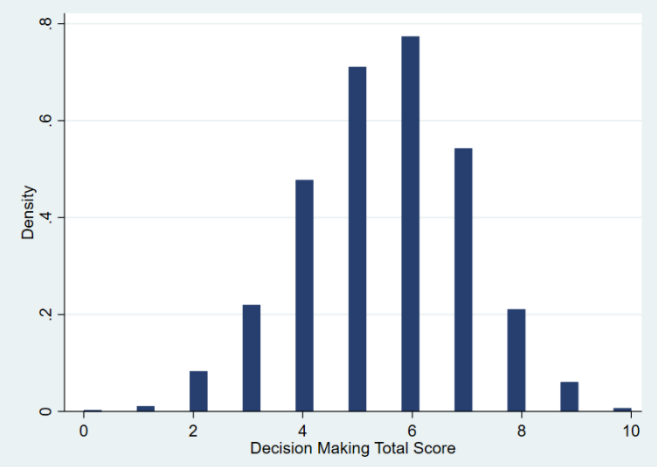


Figure 4

Distribution of ALSA Leadership Scores

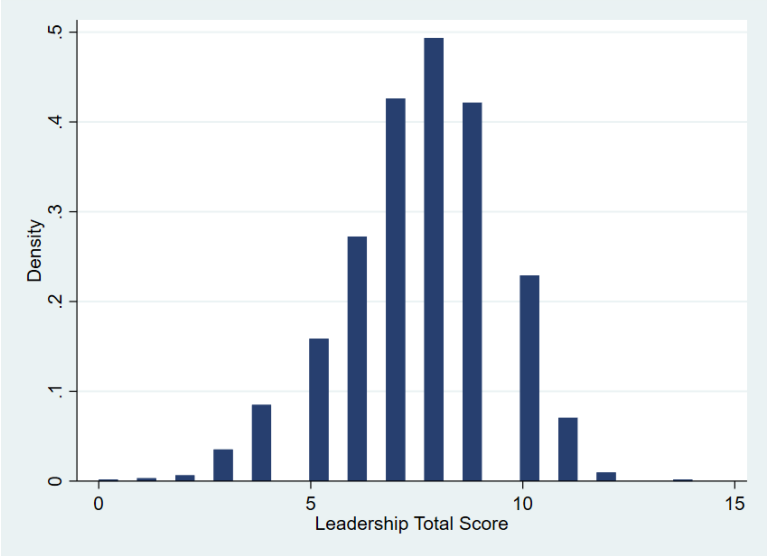


Figure 5

Distribution of ALSA Emotional Resilience Scores

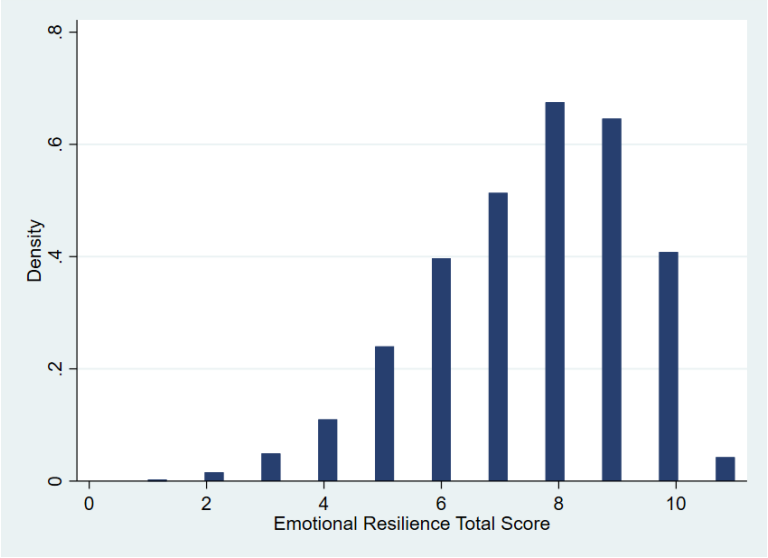


Figure 6

Distribution of ALSA Gender Knowledge and Attitudes Scores

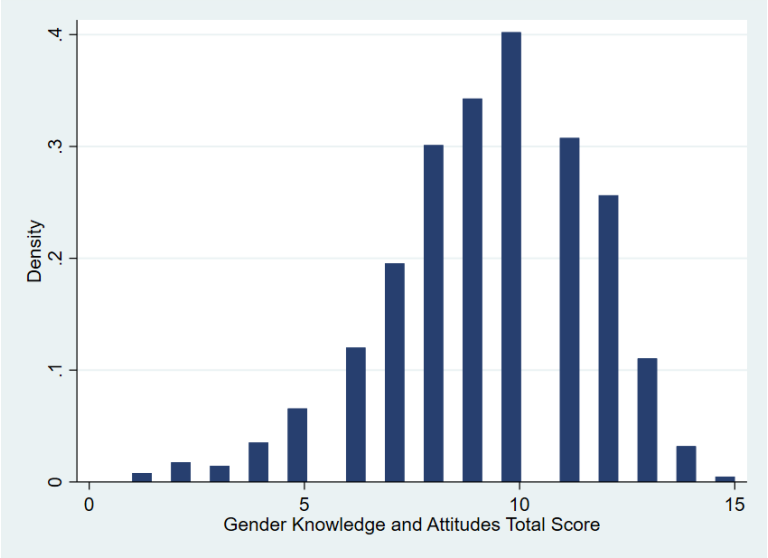
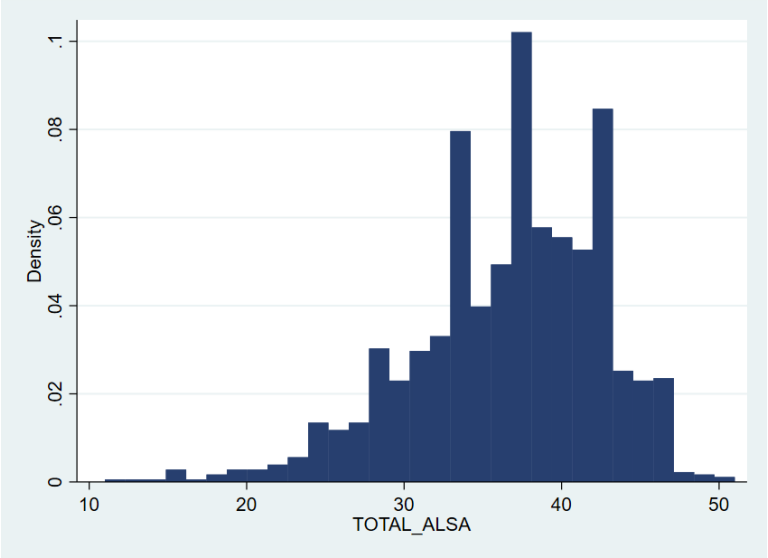


Figure 7

Distribution of ALSA Total Scores

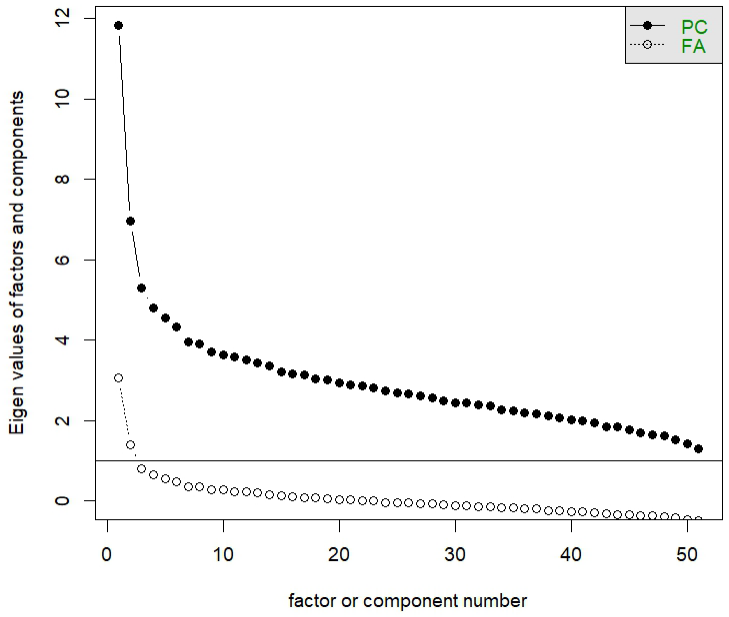


A.2. Dimensionality and factor structure

An exploratory factor analysis (using polychoric correlations) revealed that there is significant multidimensionality underlying the ALSA data. Taking all the items into account, there are three strong factors and up to three "secondary" factors. As revealed by the Scree Plot below, starting on the seventh factor, the contribution of each additional factor flattens out.

Figure 8

EFA Scree Plot



When applying a Confirmatory Factor Analysis on the data (parcels instead of individual items), focusing on five (and not six) factors, the different domains conform to the intended structure to a relatively high degree. Coefficients are significant and covariances between related factors appear significant as well (see Table 6). Fit indexes are acceptable (see Table 7), yet not great: for example, TLI should be higher than 0.95 and the RMSEA should be smaller than 0.05 (see Brown, 2015, for specific recommendations on fit indexes).

Table 6

CFA Coefficients

Coefficients	Estimate	Standard Error	Z Value	P > |Z|
Collaboration	1.00			
Decision-making	0.82	0.08	10.74	0.00
Emotional Resilience	1.39	0.10	14.41	0.00
Leadership	1.33	0.16	8.56	0.00
Gender Knowledge and Attitudes	2.64	0.29	8.98	0.00
Significant covariances	**Estimate**	**Standard Error**	**Z Value**	**P > |Z|**
Collaboration and Emotional Resilience	0.41	0.10	4.30	0.00
Leadership and Gender KA	-1.09	0.19	-5.86	0.00

Table 7

CFA Fit Indexes

Fit Indexes - Full Model	Coefficient
CFI	0.95
TLI	0.84
RSMEA	0.11

A.3. Reliability considerations

Reliability coefficients were conducted using single data coefficients and test-retest methods. For single coefficients, we used coefficient H. As explained in McNeish (2017), assumptions behind coefficient Alpha are often violated. We believe that the ALSA data violates principles of tau-equivalence, and that coefficient H is more appropriate; see Table 8 for results. The interpretation of coefficient H is identical to that of Alpha. In addition, test-retest coefficients[6] were estimated on a sample of 123 students; these are also reported in Table 9.

[6] Test-retest coefficients were estimated using IRT scores

Table 8

Reliability Coefficients

Domain	Coefficient H	Test/Retest (IRT Scores)
Collaboration	**0.45**	**0.53**
Decision-making	**0.50**	**0.65**
Emotional Resilience	**0.54**	**0.74**
Leadership	**0.47**	**0.62**
Gender Knowledge and Attitudes	**0.59**	**0.65**
ALSA Total Score	**0.76**	**0.70**

Test-retest methods are closer to real reliability estimates, and we take these coefficients as the main source of information about the consistency of ALSA scores. Based on these results, at the individual level, only Emotional Resilience shows acceptable levels of reliability (0.7 or higher) and so does the total ALSA scores. The lowest reliability coefficients are observed for the Collaboration domain, below acceptable thresholds using any rule of thumb. A closer analysis on the items suggest that participants may have used high levels of guessing to respond to some items, particularly on the collaboration and decision-making domains. It is expected that guessing levels will decrease by endline, after acquiring the related content, and that reliability scores will increase. However, other measures may be taken to ensure that endline scores are reliable at the domain level, such as adding extra items and/or reconsidering scoring methodologies.

A.4. Information functions

An alternative way to think about the reliability of scores is by examining information functions. Information functions show the amount of meaningful information gathered by a set of items on different parts of the IRT ability scale[7]. In the

[7] Often runs from -4 to 4 and is centered at 0.

case of ALSA, we care about the information gathered at the mean, around zero. Typically, information levels of 10 or more are considered ideal (see Hambleton et al., 1991). As shown in figures 9-13, this level of information would only be reached for the ALSA total score (which is obtained by adding up the information from each construct).

Figure 9

Information Function for Collaboration Domain

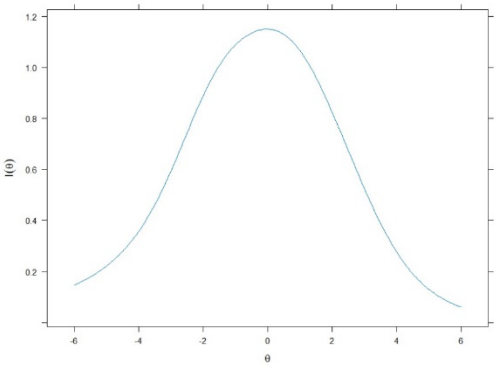


Figure 10

Information Function for Emotional Resilience Domain

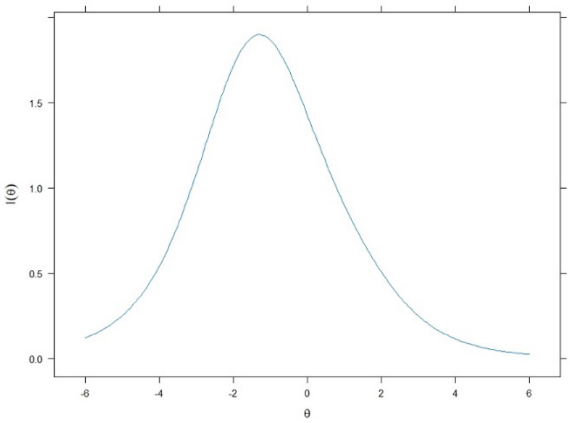


Figure 11

Information Function for the Leadership Domain

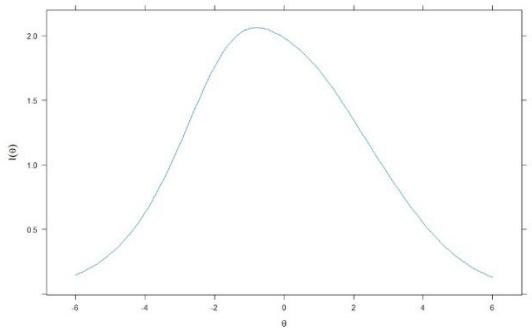


Figure 12

Information Function for the Decision-Making Domain

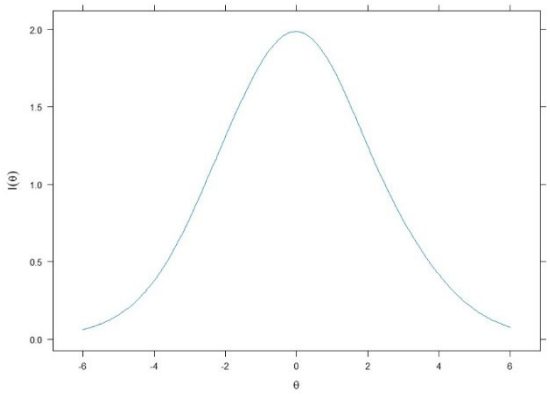


Figure 13

Information Function for the Gender Knowledge and Attitudes Domain