



Quality Assessment and Support System (QASS) Measure | Validation Study

FINAL STATISTICAL REPORT

04 December 2024



DATE

04 December 2024

CONTACT INFORMATION

Genesis Analytics (Pty) Ltd South Africa

Office 3, 50 Sixth Road Hyde Park, **Johannesburg**

Genesis Analytics Ltd Kenya

9th Floor Europa Towers, Lantana Road, Westlands, **Nairobi**

Genesis Analytics Côte d'Ivoire

Royal Work Club, Immeuble Arc-En-Ciel, Avenue Chardy, **Abidjan**

Genesis Analytics United Kingdom

Office 4.06, 4th Floor, 88 Kingsway, Holborn, London, WC2B 6AA, **London**

www.genesis-analytics.com

AUTHORS

Dr. Jürgen Becker

Dr. Michelle Henry

Ms. Georgi Borros

EDITORS

Dr. Jan Schenk

CONTACT PERSON

Dr. Jan Schenk

jans@genesis-analytics.com

Table of Contents

Table of Contents	3
1 Introduction	4
Data description	5
Statistical Modelling	6
Section A Sample Characteristics	10
Section B Data Integrity - Descriptives	15
QASS	15
ECERS	21
ELOM	23
Section C Inter-rater Agreement	27
Section E Reliability	31
Section F Convergent Validity	34
Section G Criterion Validity	36
QASS Main Effects	36
Multiple regression analysis	39
Conclusion	43
Recommendations	43
Statistical recommendations	43
Use case recommendations	44
Appendix	45
A. Demographics broken down by ELP fee-band	45
B. Boxplots of QASS, ECERS & ELOM Domain and Sub-domains by ELP fee-band	46
C. CFA at QASS Domain Level	53
D. Correlation Heatmaps	55
REFERENCES	60

1 | Introduction

The **statistical validation** of the **Quality Assessment and Support System (QASS)** measure was undertaken by AX group in partnership with Genesis Analytics, on behalf of Ilifa Labantwana. The QASS, developed in collaboration with the Department of Basic Education in South Africa, aims to measure the quality of national Early Learning Programmes (ELPs).

The QASS assesses the quality of ELP's across five key domains:

1. Learning Programme
2. Programme Management and Leadership
3. Parent and Community Engagement
4. Inclusiveness
5. Health and Safety

The QASS is designed to be completed by assessors, who provide ratings based on a combination of playroom observations, programme structure evaluations, practitioner interviews, and principal interviews. Ratings are provided across domain-specific indicators on a scale corresponding to the programme evaluation outcomes representing ELP quality. The QASS takes a similar form to Behaviourally Anchored Rating Scales (BARS), in which response options for assessors are tied to specific and clearly defined behaviours, selected by subject matter experts as indicators of the focal quality or competence. Psychometric validation, our key objective in this report, involves establishing the **reliability (consistency) and validity (veracity)** of a measurement instrument.

Table 1 below provides a detailed illustration of the 53 items making up the 5 domains of the QASS.

Table 1. QASS Structure

ELP DOMAIN	ELP SUBDOMAIN	Observation Methods	Number of items	Item number
DOMAIN 1				
Learning Programme (22 Items)	<i>Equipment, Play method, & Layout (Subdomain 1A)</i>	Playroom observation	4	1-4
	<i>Daily Program (Subdomain 1B)</i>	Playroom observation	4	5-8
	<i>Activities (Subdomain 1B)</i>	Playroom observation	2	9-10
	<i>Interaction with Children (Subdomain 1C)</i>	Playroom observation, Programme	4	11-14
	<i>Teaching Strategies (Subdomain 1C)</i>	Playroom observation	5	15-18
	<i>Curriculum (Planned Experiences, Routines) (Subdomain 1D)</i>	Playroom observation	1	19
	<i>Planning, Observing, & Assessing (Subdomain 1D)</i>	Practitioner interview	3	20-22
DOMAIN 2				
Leadership & Management (11 Items)	<i>Staffing Qualifications (subdomain 2A)</i>	Principal Interview	2	23-24
	<i>Professional Development (subdomain 2A)</i>	Principal Interview	2	25, 29
		Practitioner Interview	3	26-28, 30
	<i>Administration and Governance (subdomain 2B)</i>	Principal Interview	4	31-34

DOMAIN 3				
Parent and Community Engagement (10 items)	<i>Orientation</i>	Principal Interview	1	35
	<i>Communication and Reporting</i>	Principal Interview	3	36-38
	<i>Involvement with Programme</i>	Principal Interview	2	39-40
	<i>Parent Education</i>	Principal Interview	2	41-42
	<i>Community Involvement</i>	Principal Interview	2	43a, b
DOMAIN 4				
Inclusiveness (6 items)	<i>Disability, Chronic Illness</i>	Principal Interview	4	44-47
	<i>Language of Learning and Teaching</i>	Practitioner Interview	1	48
	<i>Inclusivity</i>	Playroom observation	1	49
DOMAIN 5				
Nutrition & Safety (4 items)	<i>Nutrition</i>	Principal Interview	1	50
		Playroom observation		50
	<i>Safety</i>	Principal Interview	3	51-53

Data description

In total, **data was collected for 288 ELPs**. In some instances, two raters completed the QASS to ascertain inter-rater reliability. The QASS was completed by **two independent raters for 234 ELPs**. An important metric to consider in terms of reliability for rater-based assessments specifically is inter-rater reliability (Babbie, 2014). This metric is mathematically very similar to Cronbach's alpha, but rather than consistency between item-level measurements across cases, it assesses consistency of assessor ratings. Usually, inter-rater reliability requires that multiple assessors evaluate the same case. In simpler terms, if many assessors provide similar QASS ratings for the same ELP, reliability of the QASS is supported.

Data was also obtained for the **ECERS (n = 87 ELPs)** and **ELOM (n = 271 ELPs)** to establish convergent and criterion-related validity. It should be noted that the raters who completed the QASS were different to those who completed the ECERS. This was done to ensure the independence of the ratings for the QASS and ECERS.

ECERS

The Early Childhood Environment Rating Scale (ECERS) is a tool used to assess the quality of ELP environments. For this assessment, a combination of the ECERS-E, ECERS-R and ECERS-3 was utilised. The ECERS-3 evaluates the quality of a classroom by observing the environment and interactions between adults and children. It assesses the needs of children in areas such as cognitive, social-emotional, physical, health, and safety. The ECERS-R measures the quality of group programs for infants and toddlers. The ECERS-E is designed to be used with the ECERS-R, and measures quality in literacy, mathematics, science and environment, and diversity practices in the preschool setting.

ELOM

The ELOM Assessment is a standardised tool that measures performance across five key developmental domains for children aged 50 to 59 months and 60 to 69 months:

1. Gross Motor Development (child's ability to control the large muscles of the body).
2. Fine Motor Coordination and Visual Motor Integration (child's ability to control the small muscles and coordinate small movements with visual information perceived by the eyes).

3. Emergent Numeracy and Mathematics (early maths skills such as the ability to understand number concepts, symbols, shapes and size).
4. Cognition and Executive Functioning (child's ability to stay focused, think critically, stay focused, form concepts, attend to instructions, and control impulses).
5. Emergent Literacy and Language (child's ability to communicate effectively. This includes learning to speak in full sentences, recognise the initial sound of words, name common objects, relay events and listen to and understand stories told to them).

For each of the ELOM domains, and for learning overall, the tool helps determine whether children are:

- On track for achieving the expected standard for their age
- Falling behind the standard
- Falling far behind and in need of significant assistance to reach the standard

ELOM results can be used to guide programme enhancement, to optimise resource allocation, to compare the relative effectiveness of different types of programmes, to assess the impact of risk factors on young children, to inform systems enhancements and to enable population-level surveillance of child outcomes. This tool is available in all of South Africa's 11 official languages. It is designed to be used by an accredited ELOM assessor or trained professional psychometrist e.g. occupational therapist. ELOM data was collected from 4 children at each ELP and then we took the aggregate score per ELP from that.

Statistical Modelling

The data analysis plan is presented in Table 2 below.

Table 2. Data analysis plan

Imperative	Analytic Method	Data Requirement	Criteria / Guideline
Data Integrity - Descriptives	Bar Plots	N / A	Should demonstrate discrimination and spread
Data Integrity - Descriptives	Skewness & Kurtosis	Response levels at least 5	Skewness > 1 or < -1 Kurtosis < 2 or > 4
Inter-rater Agreement	Cohen's Kappa	Compare ELP ratings for the same program across multiple raters	Kappa > 0.4
Construct Validity	Confirmatory Factor Analysis	At least 5 response levels Min 4 indicators per subdomain	Model fit criteria confirmed p (Chi square) < 0.05, RMSEA < 0.08 CFI & TLI > 0.9 Standardised factor loadings > 0.30
Reliability	Cronbach's alpha		> 0.7
Convergent Validity	Correlations (ECERS)	N / A	r > 0.5
Criterion-Related Validity	Linear Regression (ELOM scores ~ QASS)	N / A	p (F statistic) < 0.05 R squared shows notable (e.g. > 0.2) variance accounted for

A preliminary evaluation of central tendency within the data was undertaken to establish the appropriacy of data for modelling. We investigated to what degree data distributions were normal, skewed, or kurtotic.

In terms of reliability, the most common measure is Cronbach's alpha statistic, which speaks to the internal consistency of items on the assessment. In other words, whether all QASS indicators consistently correspond to the determination of overall ELP quality when considered across ELP's. Inter-rater reliability involves assessing the same ELP across multiple raters, to determine if responses on the QASS differ between assessors despite an absence of actual ELP differences. Preferably, assessors should be diverse in terms of their demographic groupings (e.g. gender, ethnicity) to control for cross-group sources of unreliability.

Initially we decided not to investigate construct validity due to the large sample sizes needed. However, we decided to conduct Confirmatory Factor Analyses on the data, but the results should be interpreted with caution due to the relatively small sample size. Construct validity is a critical consideration for psychometric validation. The most common method for assessing construct validity is to derive a statistical model that appropriately characterises the theoretical domain(s) of an assessment, and testing whether it fits with variance in observed scores. A confirmatory factor analysis will be utilised. Although construct validity did not form part of the scope of the current validation effort – we decided to include the CFA analyses to provide further insights into the internal structure of the measure.

Convergent validity is concerned with the degree of correspondence between scores for two assessments (or assessment subdomains) purporting to measure the same construct. If scores (or subscores) for the QASS strongly correlate with scores on established measures of ELP quality (e.g. ECERS), construct validity is supported. This was achieved by assessing the same ELP with ECERS, and assessing score correlations.

Ordinarily, criterion-correlations are (when available) a gold standard for establishing the accuracy of quality evaluations. A 'criterion' refers to some tangible, observable, and measurable metric which directly reflects the construct under investigation – in this case, ELP quality (Babbie, 2014). In the case of the QASS, criterion was scored on the ELOM. We used linear regression models, with ELOM regressed on QASS domain scores as well as the QASS total score. Such a model enabled us to quantify the amount of variance in the criterion variable that can be explained by variance in QASS scores in terms of the R squared statistic (R^2). Substantial R^2 values indicate that criterion-related validity is met. For fairness, if no additional variance is explained when ELP demographic features (such as fee-band) are considered within the statistical model, standards for fairness are met. This can be achieved through hierarchical regression modelling, wherein r squared values are compared between the model used for criterion validity and a second model including demographic variables. If the model including demographic variables is a) statistically significant on an F-test and b) noticeably higher on R^2 , issues in fairness are indicated.

Table 3 presents a breakdown of the analytic definitions and standards we used when interpreting the results.

Table 3. Analytic definitions and standards

Descriptives	
Skewness	A measure of the symmetry of a distribution. Values over 0.5 (right skew) or under -0.5 (left skew) indicate moderate skewing of the distribution and separation of the mean, median, and mode. Values over 1 or under -1 indicate a highly skewed distribution.
Kurtosis	A measure of the peakedness or dispersion within a distribution. A value of 3 indicates a dispersion is consistent with a normal distribution. Values less than 3 indicate a widely dispersed distribution, and values greater than 3 indicate a highly peaked distribution.

Reliability (Kline, 2000)	
Cronbach's Alpha (α)	Generally, a value of $\alpha = 0.7$ and up indicates acceptable scale reliability. To assess reliability on the item level, a statistic is calculated with each item excluded sequentially. If the reliability (α) of the scale improves when an item is excluded, that item is detracting from the internal consistency of the scale
Cohen's Kappa (κ)	Measures the level of agreement between two raters or observers over and above chance. Ranges between 0 and 1 where 0 indicates a level of agreement equivalent to that expected by chance, and a value of 1 indicates perfect agreement between raters.
CFA (Barrett, 2007; Hu & Bentler, 1999; Tavakol & Wetzel, 2020).	
Chi-square (χ^2)	Tests the fit of observed data to a specified structural model. Smaller values with nonsignificant p-values are considered indicative of model fit. However, this test is considered highly sensitive and often shows misfit for generally well-fitting models tested in larger samples or with complex factor structures. For this reason, fit indices such as RMSEA, CFI, and TLI are usually considered more important for assessing CFA model fit.
Factor Loadings	Represent the strength and direction of the relationship between an observed variable and an underlying latent factor. Raw factor loadings are influenced by the measurement scale of the observed variables, meaning they are not always easily interpreted or compared.
Standardized Loadings	Factor loadings transformed to a standardized, universally comparable scale. Standardized loadings above 0.3 are typically considered acceptable.
RMSEA	An Absolute Fit Index where a value of 0 indicates a perfect model. Values closer to 0 indicate better model fit. Values below 0.05 are preferred, while values below 0.08 are acceptable and values up to 0.1 are considered marginal.
CFI & TLI	The Comparative Fit (CFI) and Tucker-Lewis (TLI) Indices are both fit statistics which compare the fit of a factor model to a baseline model. Values both range from 0 to 1 and are considered acceptable above 0.9 and excellent above 0.95.
Linear Regression modelling (Kline, 2000)	
R^2	The percentage of variance in one variable that can be explained by the variance in another. Standards depend on the predictor and criterion, but a value above 0.3 is typically considered desirable.

F

A statistic used to assess the statistical significance of model fit, representing the explained variance divided by the unexplained variance in a model. Can be used to assess fit of a single model or fit between two or more models when model F statistics are compared.

The report is broken down into seven main sections:

- (a) Sample characteristics
- (b) Data integrity – descriptives
- (c) Inter-rater Reliability
- (d) Construct Validity
- (e) Reliability
- (f) Convergent Validity
- (g) Criterion-related Validity

Section A | Sample Characteristics

Sample characteristics are presented in Tables 4 and 5 below.

On average ELP fees were R325¹, with 73 children enrolled (ranging from 7 to 254²) and 5 staff members (ranging from 1 to 37). On average the programme was open for 9 hours a day, providing children with 2 meals and 1 snack. Of note is that there were very few children with disabilities at the ELPs (on average, less than 2). In addition, very few children did not speak the language of teaching and learning (on average less than 1).

Table 4. Sample characteristics

	N	Missing	Mean	Median	Mode	SD	Min	Max
Fees (rands)	287	1	325.93	150	100	551.60	0	4300
Children present in area of enumerator observation	288	0	20.98	20	20	10.71	3	75
Staff	288	0	4.92	4	4	3.42	1	37
Children enrolled at ELP	285	3	72.50	67	50	45.56	7	254
Children enrolled at ELP:Staff Ratio	285	3	16.02	15	15	8.27	2.8	52.75
Number children with disabilities	235	53	1.43	1	0	2.09	0	20
Hours programmes open for each day	288	0	9.29	9	9	1.71	3	17
Number of meals	287	1	2.07	2	2	0.40	0	4
Number of snacks	287	1	1.26	1	1	0.57	0	4
Number of age groups separated into ³	257	31	3.20	3	3	1.09	2	8
Number children who don't speak LOLT*	288	0	0.80	0	0	2.64	0	22

*LOLT: Language of Learning & Teaching

Notably, there were some high enrollment ELPs that have skewed the enrollment mean - see Figures 0.1 and 0.2 for a clearer view of the distribution of facility size in terms of the number of children enrolled at the ELP and the number of children present in the area of enumerator observation. On the day of observation, there were **15 ELPs with 0-6 children** present, **122 ELPs with 7-19 children** present, **100 ELPs with 20-30 children** present and **51 ELPs with more than 30 children** present. The number of practitioners (excluding assistants) was another interesting measure to gain insight into the ELPs that were included in the study. Excluding 7 ELPs with only one staff member, there were 15 ELPs with one

¹This is largely according to the ECD Census Data collected in 2021 and may be subject to some inflation.

²We have excluded 3 outlier values from the children enrolled variable - 1,200, 472 and 380 - as these figures correspond to primary schools and we believe that enumerators captured the enrollment for the entire primary school rather than only the Early Learning Programme. This is something to clarify for future iterations of the principal interview.

³This measure refers to all ages covered by the ELP.

practitioner, 59 ELPs with two practitioners, 133 with 3-4 practitioners, 66 ELPs with 5-9 practitioners and 20 ELPs with 10 or more practitioners.

Figure 0.1 Distribution of Number of Children Enrolled at the ELP

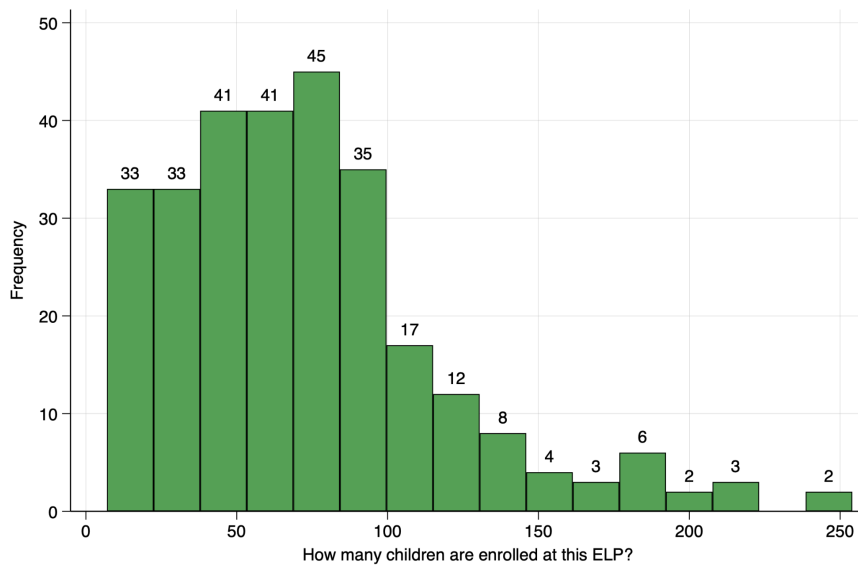
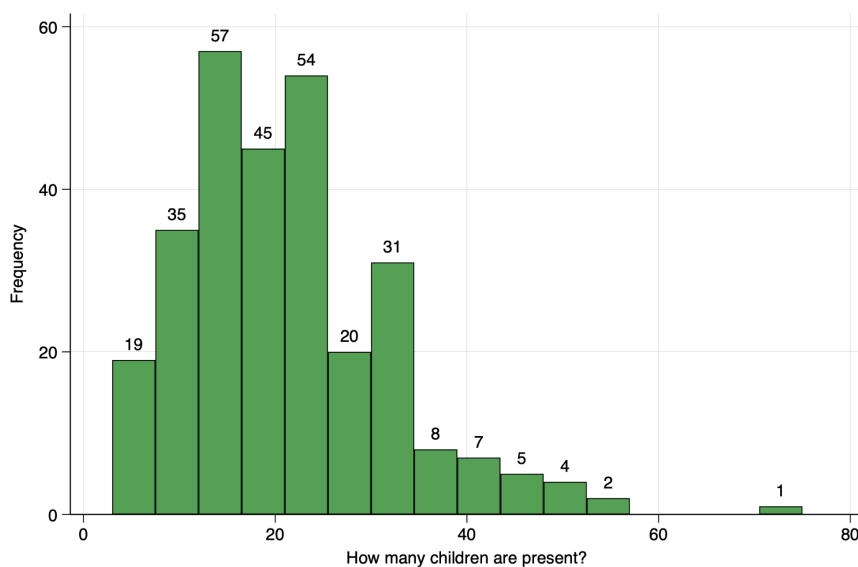


Figure 0.2 Distribution of Number of Children Present



In line with recent research conducted by *DataDrive2030* (Henry, J. and Giese, S., 2023), we use fee band as a proxy for socio-economic status and group ELPs according to the fees they charge, with Fee-band 1 charging the lowest fees and Fee-band 5 the highest. Informed by this literature, we constructed these fee-bands based on five fee-level groups as follows:

- L1: R0-R110
- L2: R111-R290
- L3: R291-R750
- L4: R751-R1750
- L5: R1751+

In this sample, the majority of ELP's were fee-band 1 and 2 ELPs (75%). Just over half the ELP's were in rural areas. Nearly all the ELP's were ECD Centres, with only 7 being playgroups. The majority were registered NPO's (83.7%) and had a separate bank account (88.5%). The majority of our data came from

Kwazulu-Natal (61.8%), and the remaining from Mpumalanga and Western Cape. The main languages of teaching and learning were Isi-Zulu (58.3%), followed by English (19.1%).

Table 5. Sample characteristics

	Counts (N)	% of Total
Facility Fee-band		
1	112	38.9%
2	104	36.1%
3	41	14.2%
4	22	7.6%
5	9	3.1%
Geo Type		
Rural	149	52.1%
Urban	137	47.9%
ELP Type		
ECD Centre	281	97.6%
Playgroup	7	2.4%
Province		
Kwazulu-Natal	178	61.8%
Mpumalanga	42	14.6%
Western cape	68	23.6%
Registration		
None of these	10	3.5%
Registered as NPO	241	83.7%
NPO application in progress	5	1.7%
Privately owned ELP	32	11.1%
Programme has separate bank account		
No	33	11.5%
Yes	255	88.5%
Children are Age Grouped		
No	31	10.8%
Yes	257	89.2%
Language of Learning & Teaching		
Afrikaans	15	5.2%
English	55	19.1%
Isi-Xhosa	19	6.6%
Isi-Zulu	168	58.3%
Siswati	31	10.8%

Given the likely disparity between ELP fee-bands in terms of resources available and the socio-economic status of children, it is important to understand how ELP fee-bands may impact QASS scores. We decided to split the data down into 2 groups: Fee bands 1-3 and Fee bands 4-5.

Sample characteristics were also broken by ELP fee-band, presented in Tables 6a and 6b below (additional demographic information broken by ELP fee-band can be found in Appendix A).

As one would expect, fees were higher for fee-band 4-5 ELPs, and these ELPs had more staff and fewer children enrolled. Fee-bands 1-3 versus fee-bands 4-5 ELPs were similar in terms of number of snacks and meals provided, hours they are open per day, and splitting children into age groups. Interestingly, a higher number of children in fee-band 4-5 ELPs did not speak the language of teaching and learning⁴.

Table 6a. Sample characteristics by ELP fee-band

	Fee Band	N	Missing	Mean	Median	SD	Min	Max
Fees (rands)	1-3	257	0	173.44	120.00	137.24	0	600
	4-5	30	1	1632.30	1275.00	928.89	800	4300
Children present	1-3	257	0	21.70	20	10.86	3	75
	4-5	31	0	15.03	14	7.13	4	35
Staff	1-3	257	0	4.71	4	3.134	1	37
	4-5	31	0	6.61	5	4.96	1	21
Number of children with disabilities	1-3	208	49	1.35	1.00	2.08	0	20
	4-5	27	4	2.04	1	2.18	0	7
Hours programme open each day	1-3	257	0	9.24	9	1.68	3	17
	4-5	31	0	9.71	10	1.95	4	12
Number of Meals	1-3	256	1	2.07	2.00	0.36	0	4
	4-5	31	0	2.09	2	0.65	0	4
Number of Snacks	1-3	256	1	1.21	1.00	0.51	0	3
	4-5	31	0	1.68	2	0.83	0	4
Number of age groups separated into	1-3	231	26	3.15	3	1.02	2	8
	4-5	26	5	3.65	3.00	1.55	2	7
Number of children who don't speak LOLT*	1-3	257	0	0.68	0	2.34	0	22
	4-5	31	0	1.81	0	4.33	0	21

Table 6b. Language of learning and teaching by ELP fee-band

Language of Learning & Teaching	Fee-band	N	%
Afrikaans	1-3	14	4.9%
	4-5	1	0.3%
English	1-3	27	9.4%
	4-5	28	9.7%

⁴It is important to note that ELPs in fee-bands 4&5 had small sample sizes (n=22 and n=9, respectively) and so sample characteristics for these groups should be interpreted tentatively.

Isi-Xhosa	1-3	19	6.6 %
	4-5	0	0.0 %
Isi-Zulu	1-3	166	57.6 %
	4-5	2	0.7 %
Siswati	1-3	31	10.8 %
	4-5	0	0.0 %

Section B | Data Integrity - Descriptives

Section B examines the suitability of the data for the application of our statistical methods. This is done using descriptive techniques that largely examine the distributional characteristics. Notably, we had 234 ELP's where 2 raters rated the QASS. The analyses presented below have therefore used the average of these 2 raters for the 234 ELP's and the remaining 54 ELP's with one rater. It should also be noted that QASS included several items where "N/A" was an option (namely items 13, 45, 48, 26, 27, 28, 30):

- Item 13: 0 N/As (Domain 1)
- Item 26, 27, 28: 20 N/As out of 522⁵ ratings (Domain 2)
- Item 30: 2 N/As out of 522 ratings (Domain 2)
- Item 45: 163 N/As out of 522 ratings (Domain 4)
- Item 48: 389 N/As out of 522 ratings (Domain 4)

In these instances, we have pro-rated these items when N/A has been selected, so as not to penalise sites where the scoring was irrelevant. In the analysis, the pro-rated items are labelled using the suffix '_pro', which is reflected in items such as 'item_45_pro'.

It is of key importance to test how scores are ultimately distributed, aiming for an approximately normal distribution. In line with having appropriately distributed results, scoring needs to have a level of variability in order to capture varying degrees of quality across ELPs.

A perfect normal distribution has a kurtosis of 0⁶ and a skewness of 0. The 'acceptable' ranges of kurtosis and skewness for an approximately normal distribution are as follows:

- Skewness: < 0.5 and > -0.5 (ideally) or < 1 and > -1 (pending impact on the mean)
- Kurtosis: < 1 and > -1

Overall, QASS domain, subdomain and total scores demonstrate a relatively normal distribution based on skewness and kurtosis, with Domain 5 bearing a slight rightward skewness, and subdomain 1c bearing a slight leftward skewness.

QASS

Domain Descriptives

QASS domain descriptives for the 288 ELPs forming part of this study are presented in Table 7 below. The measures of central tendency (mean, median) are not directly comparable due to the difference in range across dimensions. Overall, the descriptives show good variation (in terms of standard deviation (SD) and range as shown by the minimum and maximum values) and approximate normally distributed data. Domains 1, 2 and 3 in particular have excellent approximations to the normal distribution both in terms of skewness and kurtosis. Domains 4 and 5 show the most skewness relative to the normal distribution gold standard (although the skewness remains under the threshold for high concern - which would correspond to a skewness > 1 or < -1), while kurtosis values remain within the desirable range.

⁵ These are the total ratings including the inter-rater assessment at 234 of the 288 ELPs.

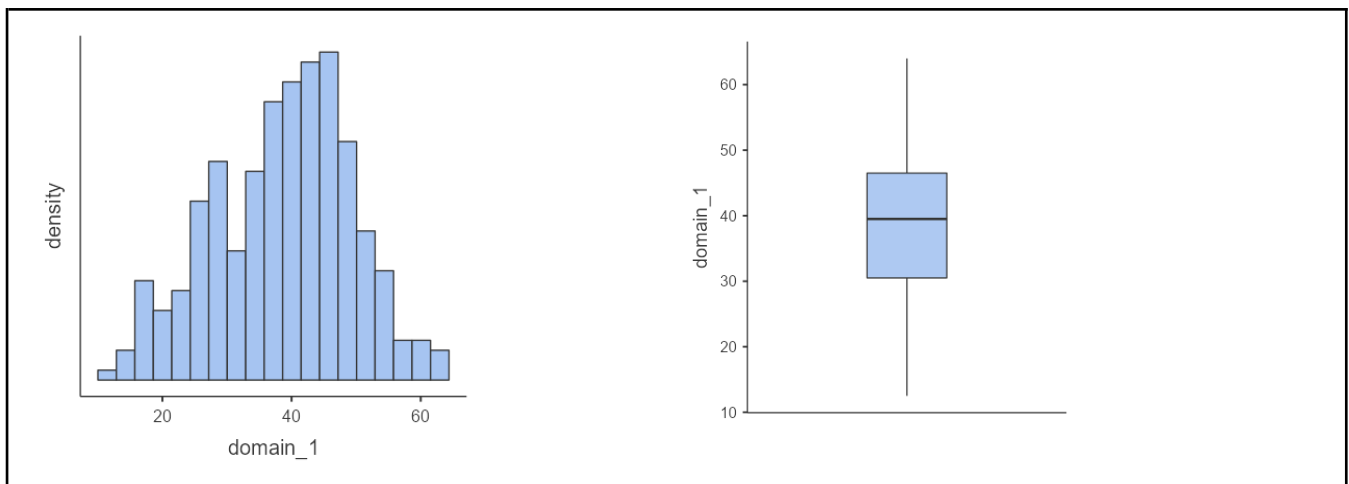
⁶ This standard for kurtosis corresponds to analysis conducted using SPSS statistical software. Other statistical software (such as Stata) use a kurtosis of 3 for a normal distribution.

Table 7: QASS domain descriptives

	Domain 1: Learning Programme	Domain 2: Leadership & Manageme nt	Domain 3: Parent & Community Engageme nt	Domain 4: Inclusivene ss	Domain 5: Nutrition & Safety	Total Score
N	288	288	288	288	288	288
Missing	0	0	0	0	0	0
Mean	38.6	17.8	14.5	10.7	4.90	86.4
Median	39.5	17.0	14.0	11.0	5.00	86.0
SD	10.6	4.89	3.49	3.11	1.76	18.6
Minimum	12.5	4.00	5.00	3.00	1.00	30.0
Maximum	64.0	30.0	23.5	17.0	11.0	131
Skewness	-0.25	0.10	0.06	-0.13	0.49	0.002
Std. error skewness	0.14	0.14	0.14	0.14	0.14	0.14
Kurtosis	-0.43	-0.01	-0.41	-0.69	0.71	-0.16
Std. error kurtosis	0.29	0.29	0.29	0.29	0.29	0.29

We depicted domain and total QASS scores with histograms and boxplots in Figures 1-6 below. The plots show the spread, shape and any outliers in the data⁷. While a few data values were labelled as outliers, we did not remove any outliers in the data analyses. These visualisations support the descriptives, with Domains 1, 2 and 3 reflecting a relatively normal shape, while Domains 4 and 5 are slightly less well defined. The QASS total score shows a good distribution, with almost 0 skewness and a mesokurtic (normal distribution) shape.

Figure 1. QASS Domain 1: Learning Programme



⁷ Outliers are any data value greater or less than 1.5 x interquartile range.

Figure 2. QASS Domain 2: Leadership & Management

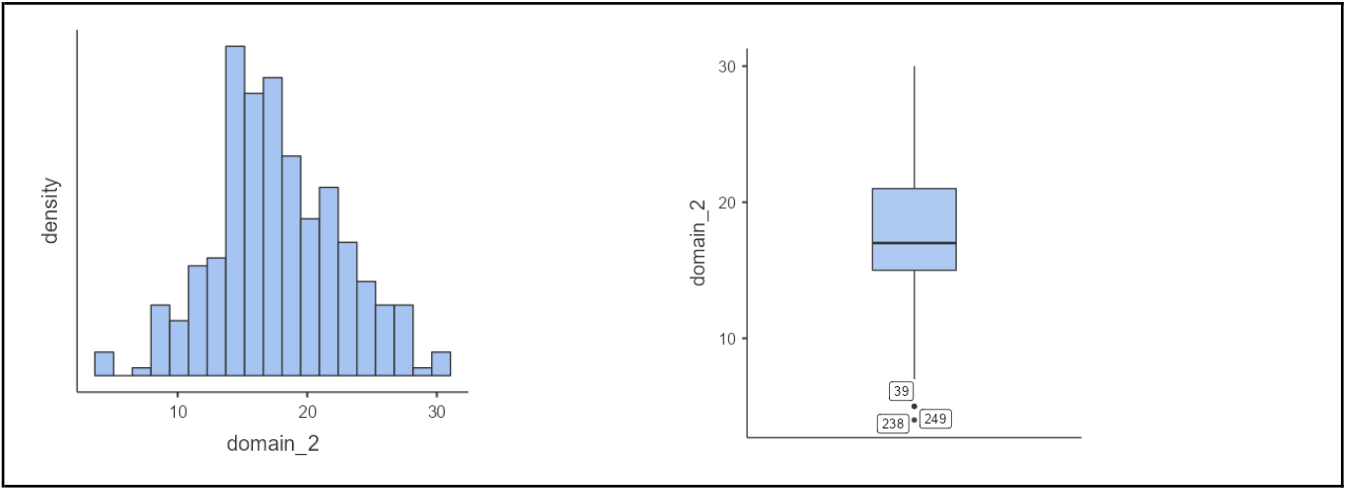


Figure 3. QASS Domain 3: Parent & Community Engagement

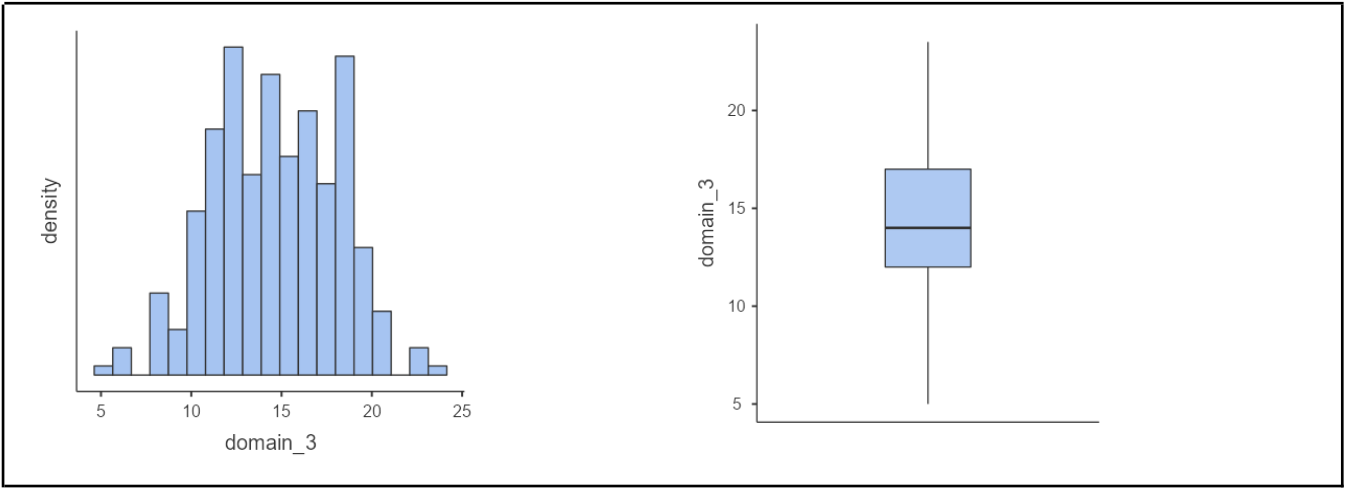


Figure 4. QASS Domain 4: Inclusiveness

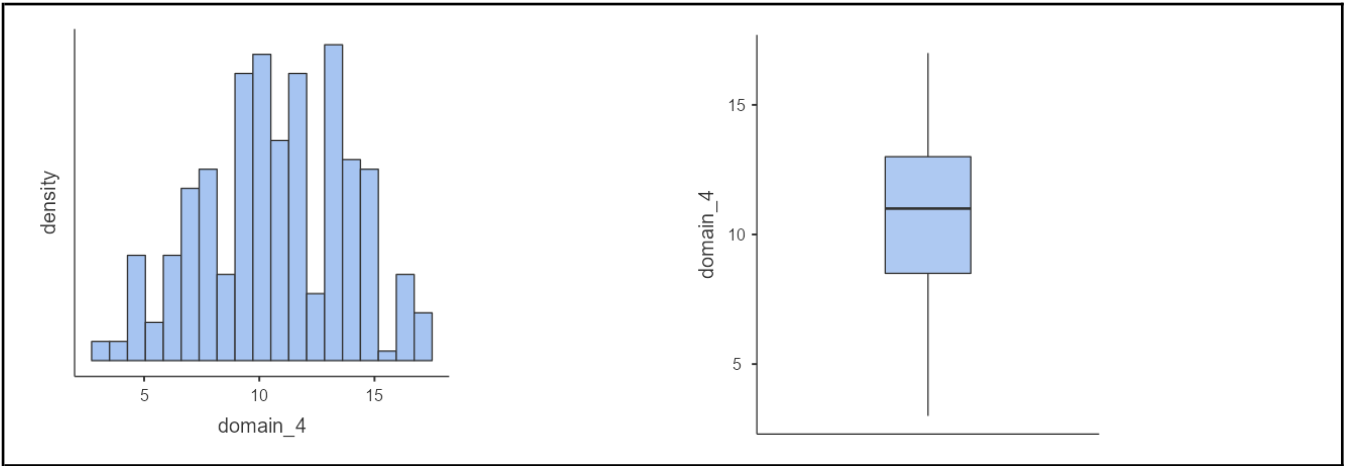


Figure 5. QASS Domain 5: Nutrition & Safety

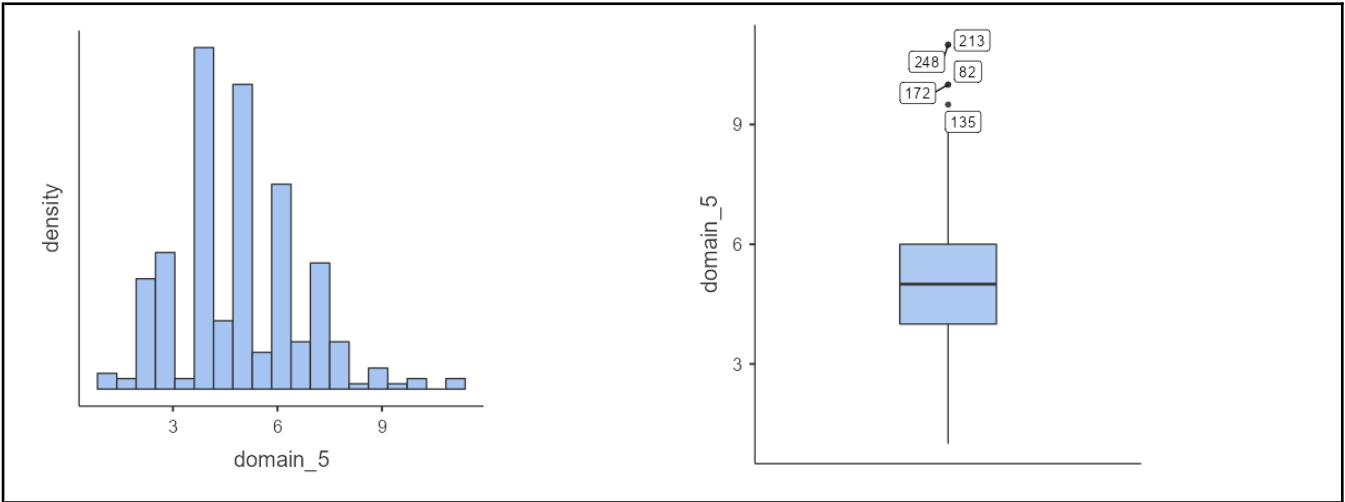
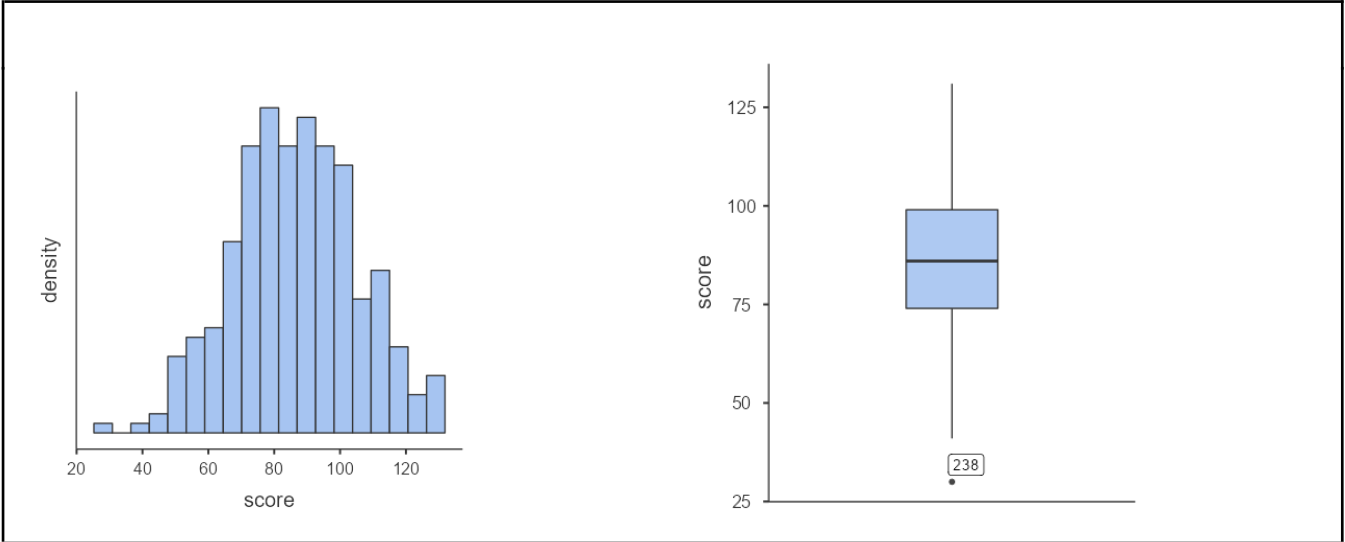


Figure 6. QASS Total Score



Subdomain Descriptives

QASS subdomain descriptives are presented in Table 8 below, and histograms and boxplots in Figures 7-12.

Table 8. QASS subdomain descriptives

	Sub 1a: Equipment	Sub 1b: Daily Programme	Sub 1c: Staff-Child Interaction	Sub 1d: Planning & Assessment	Sub 2a: Staffing	Sub 2b: Administrati on & Governance
N	288	288	288	288	288	288
Missing	0	0	0	0	0	0
Mean	7.08	9.39	15.1	7.02	11.0	6.78
Median	7.00	9.25	16.0	7.00	11.0	7.00
Standard deviation	2.25	3.39	4.72	3.06	3.23	2.36
Minimum	0.00	0.00	2.00	1.00	3.00	0.00
Maximum	12.0	18.0	24.0	12.0	20.0	12.0
Skewness	-0.24	0.13	-0.39	-0.23	0.08	-0.07
Std. error skewness	0.14	0.14	0.14	0.14	0.14	0.14
Kurtosis	0.061	-0.136	-0.528	-1.08	-0.037	-0.222
Std. error kurtosis	0.286	0.286	0.286	0.286	0.286	0.286

Figure 7. QASS Sub-domain 1a: Equipment

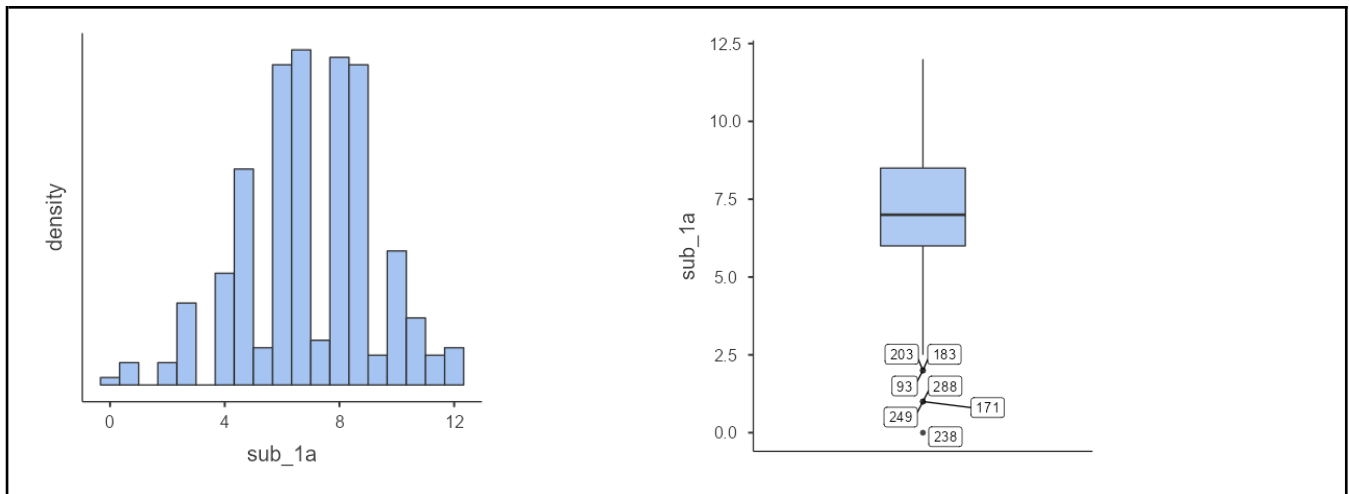


Figure 8. QASS Sub-domain 1b: Daily Programme

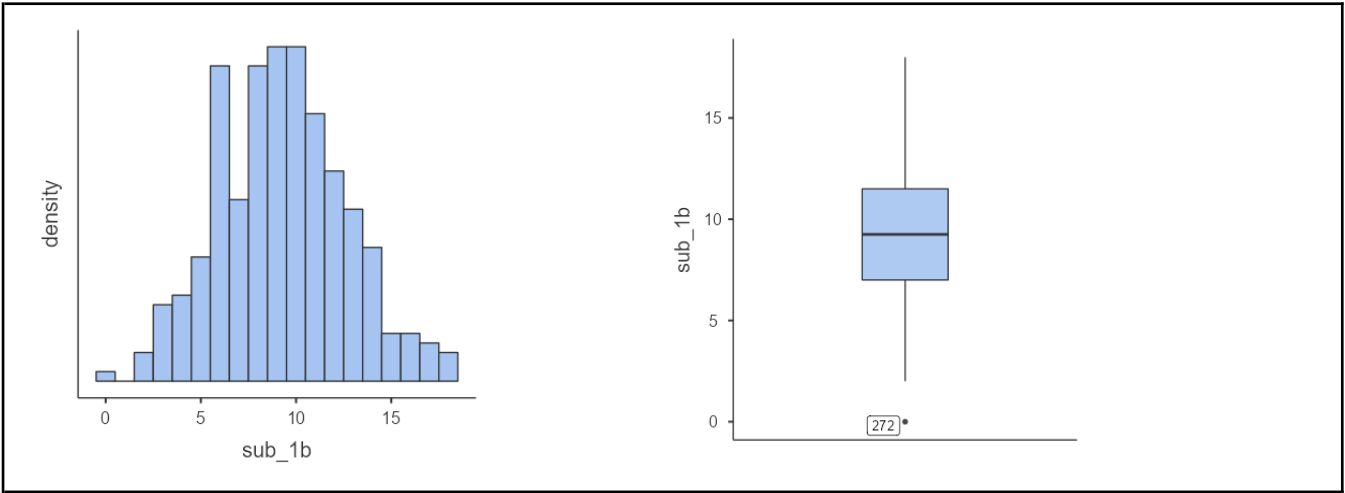


Figure 9. QASS Sub-domain 1c: Staff-Child Interaction

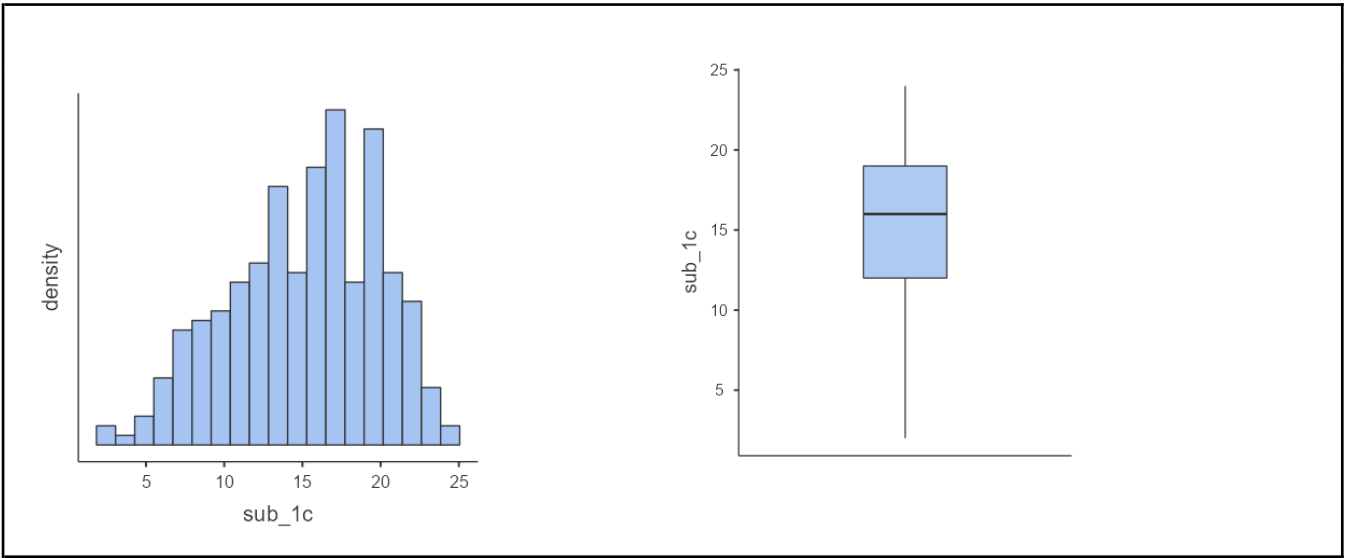


Figure 10. QASS Sub-domain 1d: Planning & Assessment

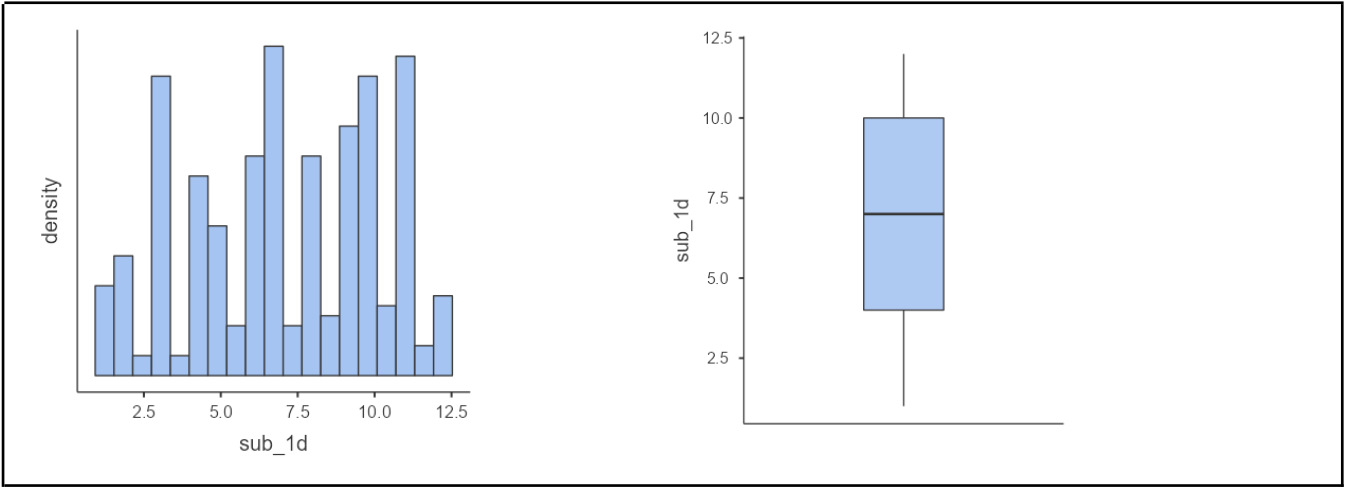


Figure 11. QASS Sub-domain 2a: Staffing

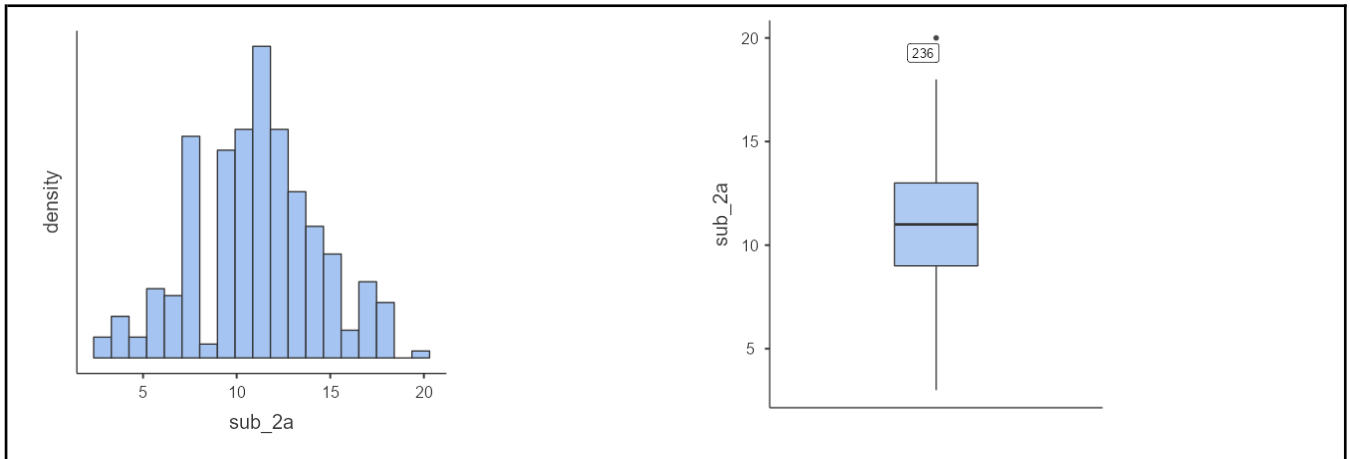
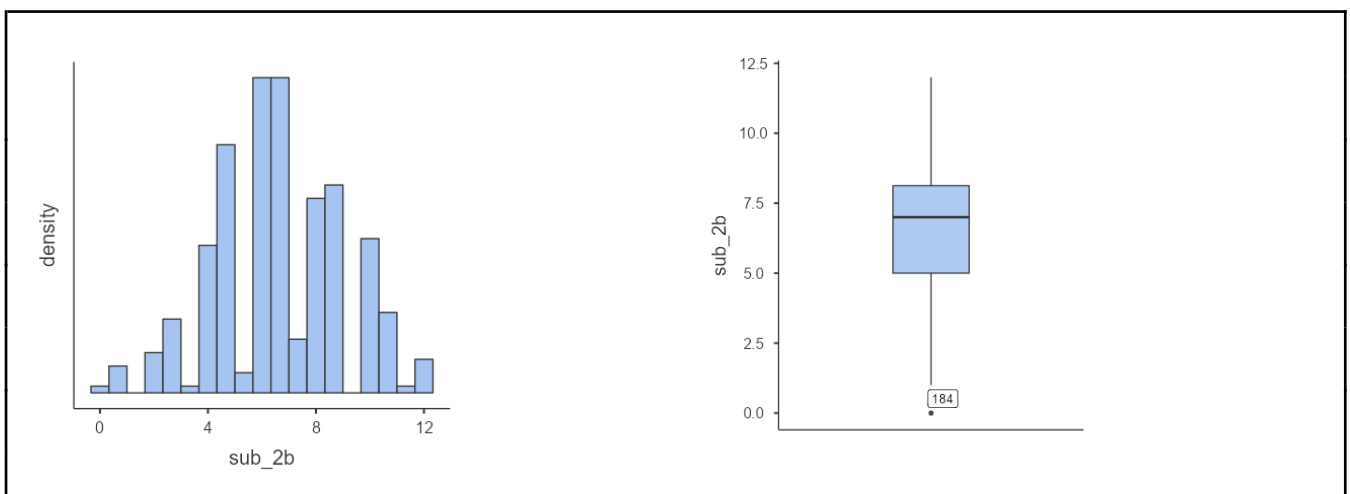


Figure 12. QASS Sub-domain 2b: Administration & Governance



ECERS

In Table 9 below the descriptive statistics of the ECERS domain are summarised. Subscale 4 (Diversity) was right skewed, and subscale 5 (Interaction) left skewed. Subscales 1 (Space & Furnishings) and 7 (Parents & Staff) have a slight left skewness. Subscale 6 (Programme Structure) was normally distributed. The mean scores are not directly comparable due to the difference in range across dimensions. Histograms and boxplots are displayed in Figures 13-17.

Table 9. ECERS descriptives

	Subscale 1: Space & Furnishings	Subscale 4: Diversity	Subscale 5: Interaction	Subscale 6: Programme Structure	Subscale 7: Parents & Staff
N	75	79	87	75	72
Missing	12	8	0	12	15
Mean	26.4	6.57	13.8	12.5	21.7
Median	26	6	15	12	20.0
Standard deviation	9.25	3.70	5.43	4.66	6.78
Minimum	9	3	3	3	11
Maximum	49	21	21	21	37
Skewness	0.39	1.49	-0.47	-0.05	0.47
Std. error skewness	0.28	0.27	0.26	0.28	0.28

Table 9. ECERS descriptives

	Subscale 1: Space & Furnishings	Subscale 4: Diversity	Subscale 5: Interaction	Subscale 6: Programme Structure	Subscale 7: Parents & Staff
Kurtosis	-0.56	2.67	-0.79	-0.82	-0.75
Std. error kurtosis	0.55	0.54	0.51	0.55	0.56

Figure 13. ECERS Subscale 1: Space & Furnishings

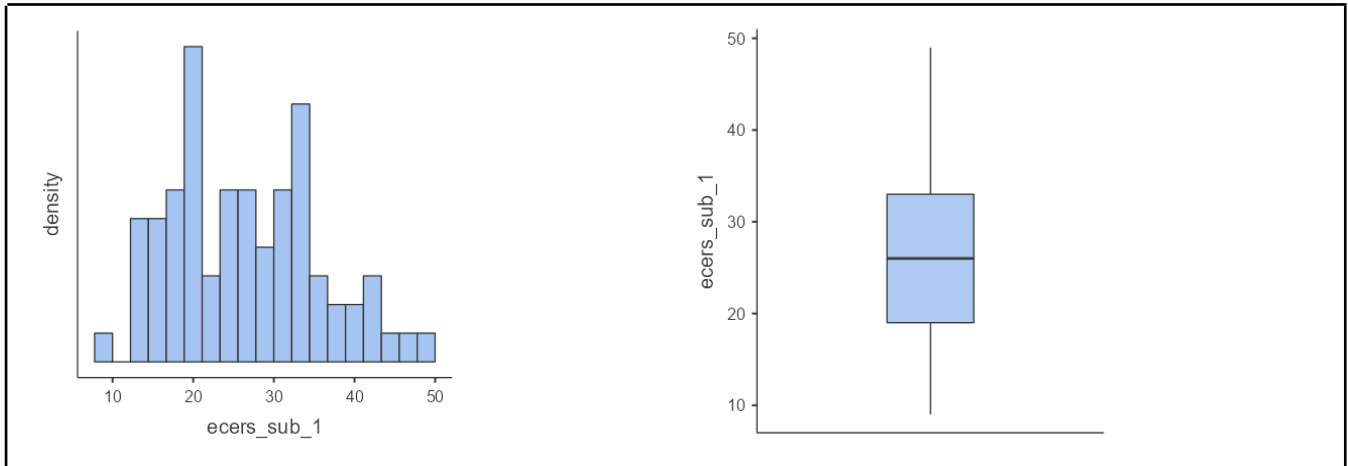


Figure 14. ECERS Diversity

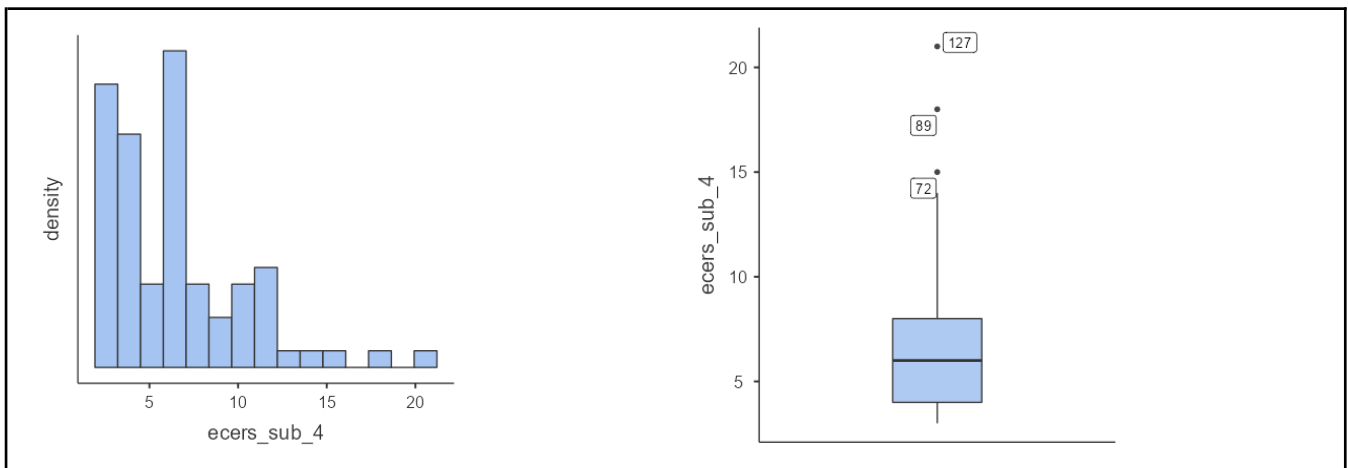


Figure 15. ECERS Subscale 5: Interaction

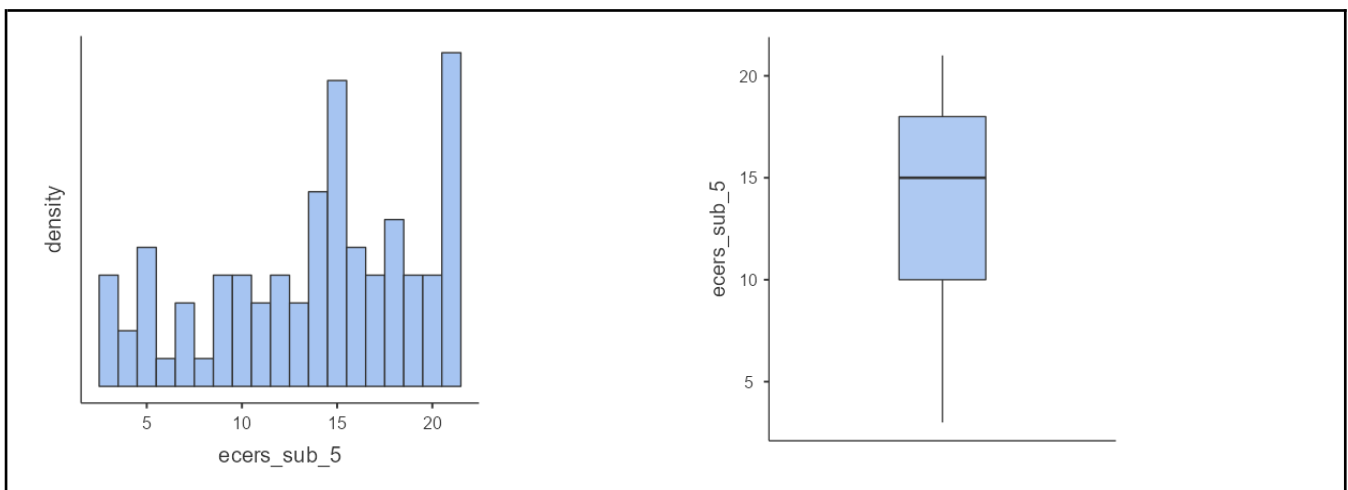


Figure 16. ECERS Subscale 6: Programme Structure

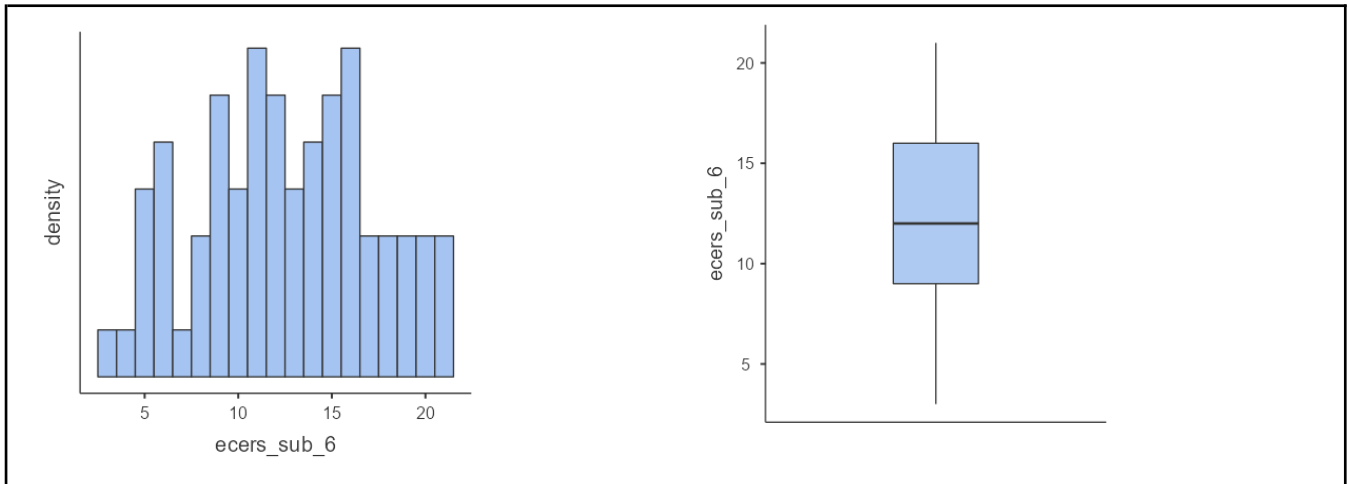
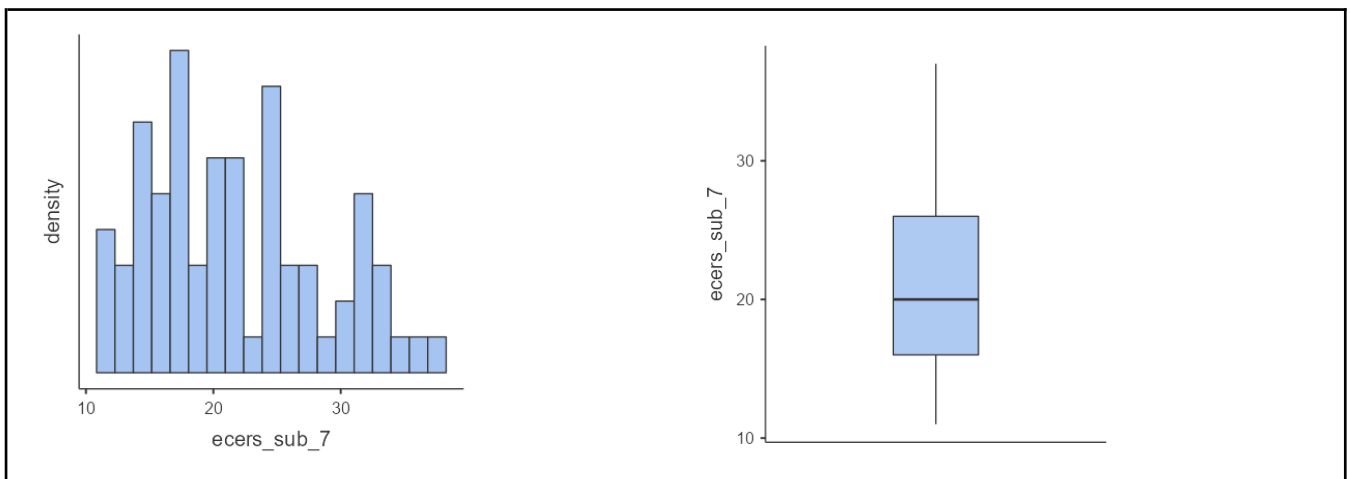


Figure 17. ECERS Subscale 7: Parents & Staff



ELOM

Table 10 below contains the most important descriptives of the ELOM measure across 5 domains. Domain 1 (Gross Motor Development), 4 (Cognition & Executive Functioning) and total scores are skewed to the right. The remaining domains are normally distributed. Histograms and boxplots are presented in Figures 18-23.

Table 10. ELOM descriptives

	Domain 1: Gross Motor Development	Domain 2: Fine Motor Coordination & Visual Motor Integration	Domain 3: Emergent Numeracy & Mathematics	Domain 4: Cognition & Executive Functioning	Domain 5: Emergent Literacy & Language	Total Score
N	271	271	271	271	271	271
Missing	0	0	0	0	0	0
Mean	8.46	10.8	8.83	7.26	11.4	46.8
Median	8.01	10.7	8.74	6.92	11.6	45.2
Standard deviation	3.41	2.66	2.82	2.88	3.14	12.0
Minimum	1.00	2.75	0.50	0.00	2.19	16.5
Maximum	20.0	19.0	18.8	18.7	18.7	93.8

Table 10. ELOM descriptives

	Domain 1: Gross Motor Development	Domain 2: Fine Motor Coordination & Visual Motor Integration	Domain 3: Emergent Numeracy & Mathematics	Domain 4: Cognition & Executive Functioning	Domain 5: Emergent Literacy & Language	Total Score
Skewness	1.03	0.30	0.46	1.04	-0.16	1.12
Std. error skewness	0.15	0.15	0.15	0.15	0.15	0.15
Kurtosis	1.82	0.412	0.79	2.35	-0.08	2.12
Std. error kurtosis	0.30	0.30	0.30	0.30	0.30	0.30

Figure 18. ELOM Domain 1: Gross Motor Development

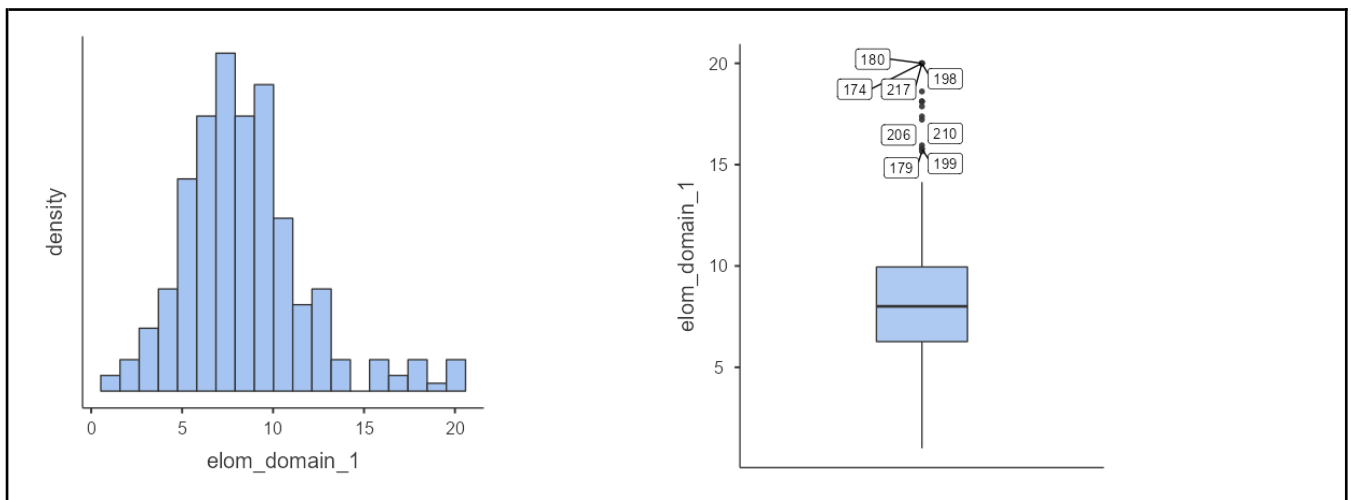


Figure 19. ELOM Domain 2: Fine Motor Coordination & Visual Motor Integration

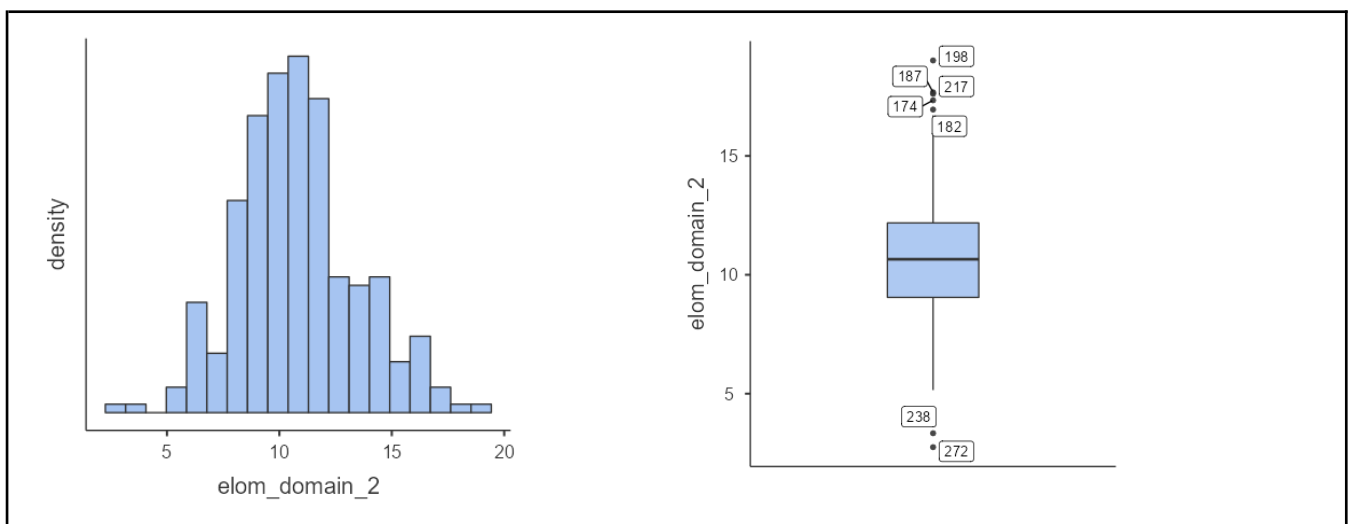


Figure 20. ELOM Domain 3: Emergent Numeracy & Mathematics

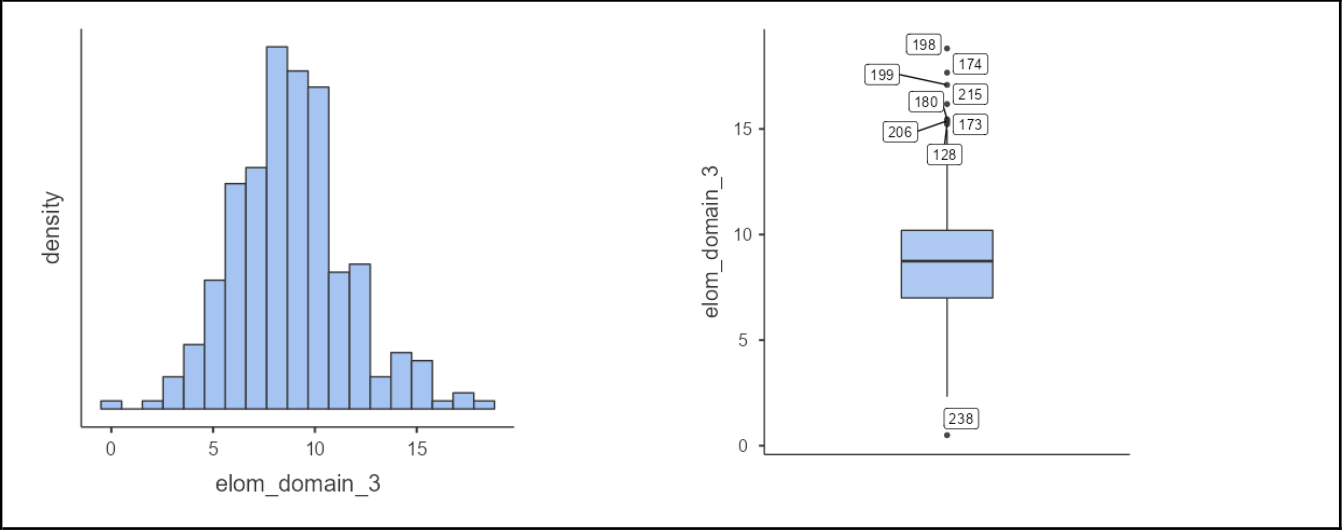


Figure 21. ELOM Domain 4: Cognition & Executive Functioning

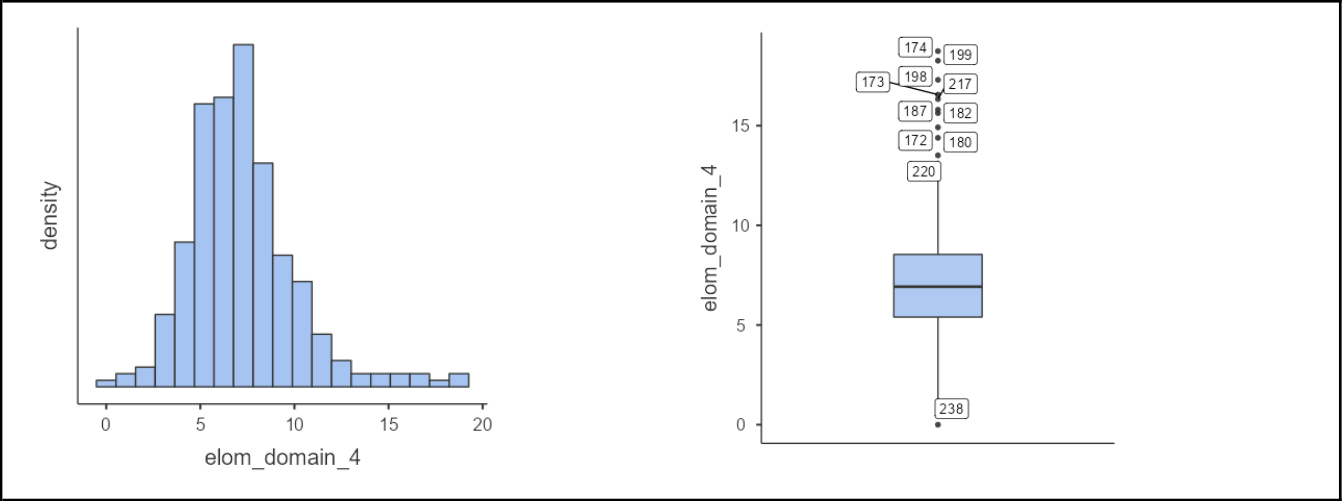


Figure 22. ELOM Domain 5: Emergent Literacy & Language

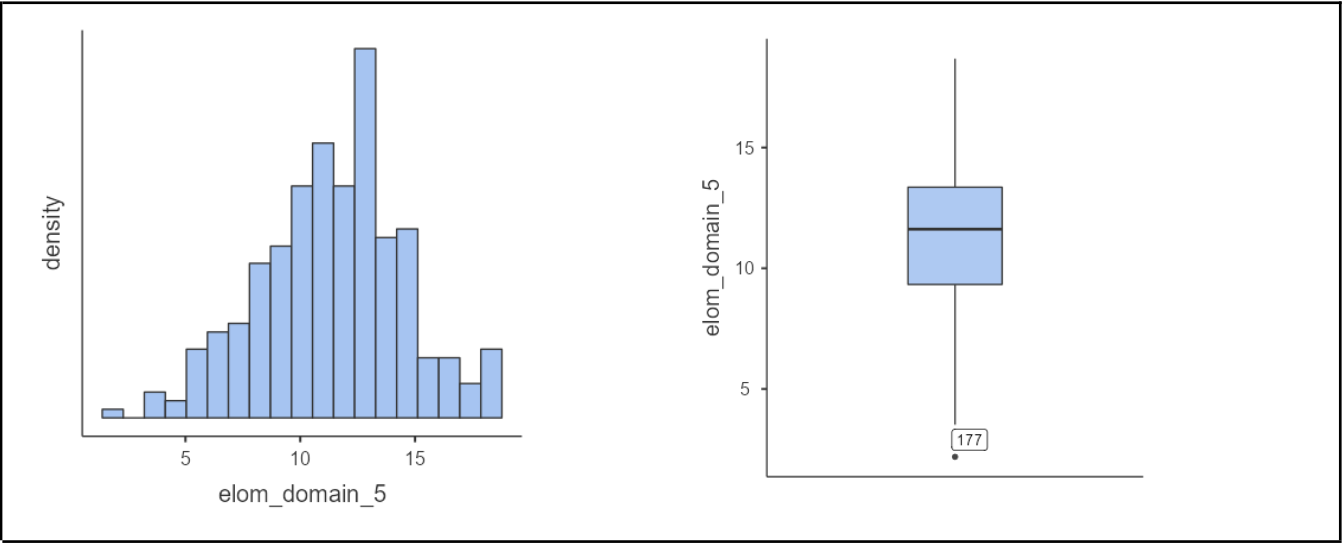
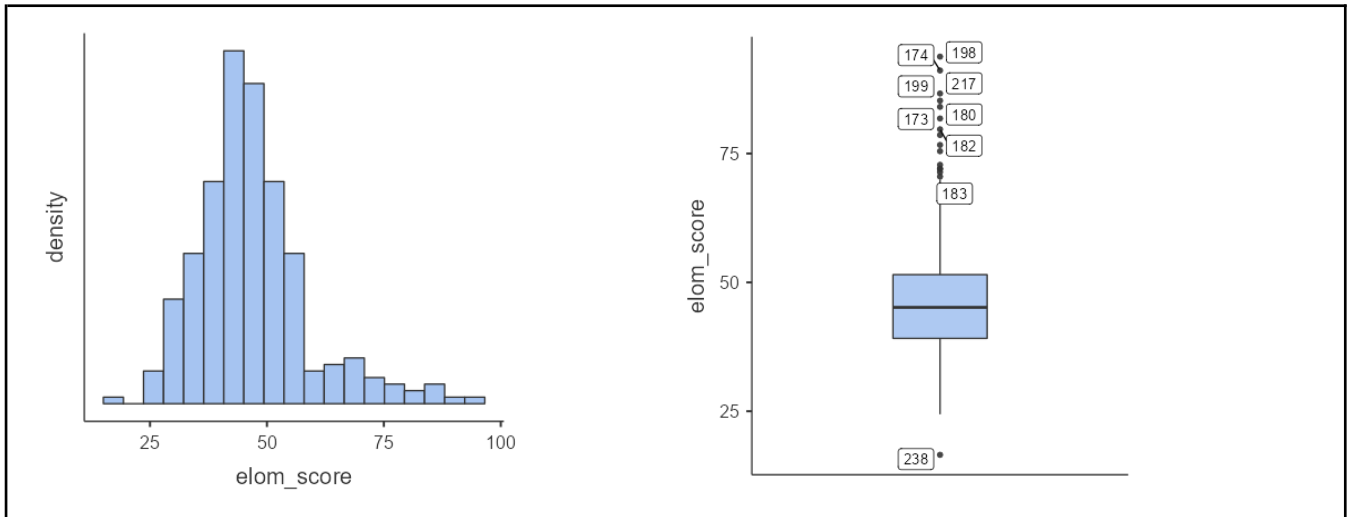


Figure 23. ELOM Total Score



QASS (Figure 24) and ELOM (Figure 25) total scores were further broken down by ELP fee-band, as depicted in the box plots below. As can be seen, fee-band 5 ELP's had a higher median QASS and ELOM score compared to fee-band 1-2 ELPs. Fee-band 3 and 4 ELPs were relatively equal in performance on the ELOM measure and also had a slightly higher median score compared to fee-band 1-2 ELPs (see Appendix B for QASS domain and subdomain scores, ECERS subdomain, and ELOM domain scores broken down by Fee-band).

Figure 24. QASS score by ELP Fee-band

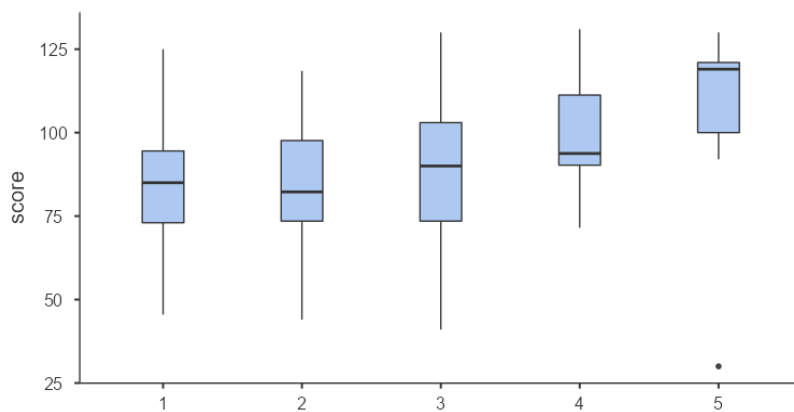
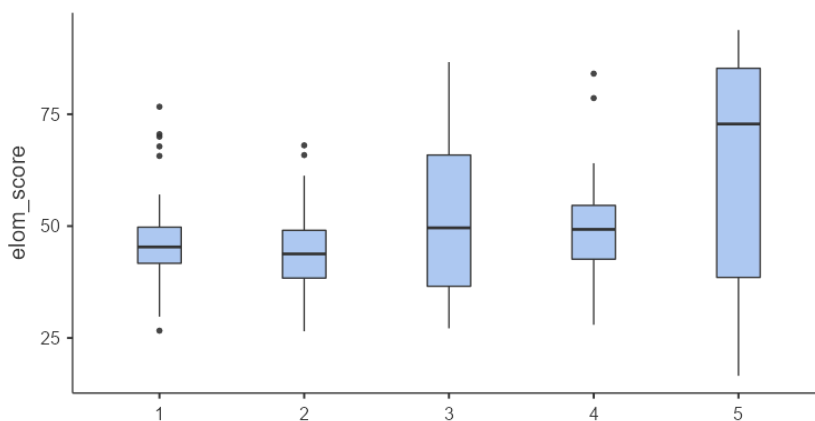


Figure 25. ELOM Score by ELP Fee-band



Section C | Inter-rater Agreement

In this section we examine inter-rater reliability using Cohen's Kappa.

Cohen's Kappa can be interpreted as follows:

None – slight	0.01 – 0.20
Fair	0.21 – 0.40
Moderate	0.41 – 0.60
Substantial	0.61 – 0.80
Almost perfect	0.81 – 1.00

Inter-rater agreement for each item in the QASS are presented in Table 11 below. As can be seen, for nearly all items there was substantial to almost perfect agreement between raters. Only two items had a moderate agreement between raters (Items 17 and 18).

The overall scale had an almost perfect agreement, with a **Kappa value of 0.833**. This indicated strong inter-rater reliability.

Table 11. QASS Inter-rater Reliability

Item 1	0.826	Item 28 (Prorated)	0.860
Item 2	0.799	Item 29	0.893
Item 3	0.764	Item 30 (Prorated)	0.943
Item 4	0.741	Item 31	0.959
Item 5	0.662	Item 32	0.934
Item 6	0.743	Item 33	0.923
Item 7	0.846	Item 34	0.914
Item 8	0.784	Item 35	0.862
Item 9	0.797	Item 36	0.909
Item 10	0.675	Item 37	0.876
Item 11	0.696	Item 38	0.886
Item 12	0.634	Item 39	0.917
Item 13	0.636	Item 40	0.916
Item 14	0.795	Item 41	0.775
Item 15	0.725	Item 42	0.858
Item 16	0.693	Item 43	0.896
Item 17	0.546	Item 44	0.899
Item 18	0.556	Item 45 (Prorated)	0.930
Item 19	0.709	Item 46	0.971
Item 20	0.932	Item 47	0.891
Item 21	0.858	Item 48 (Prorated)	0.919
Item 22	0.950	Item 49	0.713
Item 23	0.982	Item 50	0.918
Item 24	0.880	Item 51	0.948
Item 25	0.876	Item 52	0.838
Item 26 (Prorated)	0.869	Item 53	0.904
Item 27 (Prorated)	0.901	Item 43b ⁸	0.938

⁸Item 43b was included in the QASS as a test item, providing a different version of the original item 43.

Section D | Construct Validity

We ran a Confirmatory Factor Analysis on the QASS to determine whether the hypothesised model fit the data well. It is important to remember that CFA analysis on QASS violates the assumption of single source input because QASS scores are based on observations from facilitators, interviews with teachers and principals. However, CFA should provide valuable input into the factor structure of the measure. We modelled items on their relative subdomain (for Domains 1 & 2) and domain structures. We excluded Domain 4 from the analyses because it only comprised 6 items (Items 44-49), and of these, there was a zero score for 114 ELP's on Item 44, a zero score for 51 ELP's and a NA score for 88 ELP's on Item 45, and a zero score for 120 ELP's on Item 46. Due to the restricted variance in score – the items were not conducive for CFA analyses and were removed from the specified models.

The model chi-square statistic and several fit indices are commonly reported to address this question (see Tables 12 & 13). Three model fit indices that are widely applied are the root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker–Lewis index (TLI) (Xia & Yang, 2019). RMSEA is an absolute fit index, in that it assesses how far a hypothesised model is from a perfect model. On the contrary, CFI and TLI are incremental fit indices that compare the fit of a hypothesised model with that of a baseline model (i.e., a model with the worst fit). RMSEA smaller than .08 and a CFI and TLI larger than .9 indicate a relatively good model–data fit in general. Our model had a good RMSEA value (0.05), but CFI (0.78) and TLI (0.76) values are not greater than 0.9, indicating a poor model fit. The low CFI and TLI may be driven by several items whose factor loadings are below the acceptable threshold of 0.3. Another potential reason could be our limited sample size for CFA.

Table 12. CFA Test for Exact Fit

χ^2	df	p
1718	1006	< .001

Table 13. CFA Fit Measures

CFI	TLI	RMSEA	RMSEA 90% CI	
			Lower	Upper
0.778	0.761	0.0496	0.0455	0.0535

Factor loadings are presented in Table 14. Items that did not load onto their respective factors adequately (a factor loading estimate of < 0.3) are: Item 14 on subdomain 1c (Staff-Child Interaction), Item 24, 26 and 30 on Subdomain 2a (Staffing), Items 40, 42 and 43 on Domain 3 (Parent and Community Engagement), and Items 50, 51 and 53 on Domain 5 (Nutrition & Safety). These items should be flagged for review in future iterations of the measure.

We also ran a CFA using modelling of all items on the 5 broad domains of the QASS ignoring the sub-factor structure (see Appendix C).

Table 14. Factor Loadings

Factor	Indicator	Estimate	SE	Z	p
Subdomain 1a: Equipment	item_1	0.5382	0.0602	8.946	< .001
	item_2	0.3835	0.0674	5.691	< .001
	item_3	0.5088	0.046	11.07	< .001
	item_4	0.355	0.0411	8.628	< .001
Subdomain 1b: Daily Program	item_5	0.5201	0.0524	9.931	< .001
	item_6	0.5634	0.0564	9.984	< .001
	item_7	0.4628	0.0546	8.481	< .001
	item_8	0.3495	0.059	5.923	< .001
	item_9	0.4143	0.0522	7.935	< .001
	item_10	0.5143	0.0451	11.397	< .001
Subdomain 1c: Staff-Child Interaction	item_11	0.4896	0.0473	10.351	< .001
	item_12	0.5665	0.0478	11.856	< .001
	item_13_pro	0.5437	0.0684	7.954	< .001
	item_14	0.2493	0.067	3.721	< .001
	item_15	0.4992	0.0392	12.748	< .001
	item_16	0.7885	0.0627	12.577	< .001
	item_17	0.5731	0.047	12.19	< .001
	item_18	0.4959	0.0388	12.772	< .001
Subdomain 1d: Planning & Assessment	item_19	0.6865	0.0698	9.829	< .001
	item_20	0.7099	0.0638	11.126	< .001
	item_21	0.6658	0.0611	10.89	< .001
	item_22	0.5367	0.0596	9.005	< .001
Subdomain 2a: Staffing	item_23	0.3736	0.0647	5.777	< .001
	item_24	0.1428	0.0557	2.562	0.01
	item_25	0.5201	0.0511	10.172	< .001
	item_26_pro	0.0158	0.0551	0.287	0.774
	item_27_pro	0.3258	0.0499	6.526	< .001
	item_28_pro	0.4029	0.0611	6.594	< .001
	item_29	0.3459	0.034	10.176	< .001
	item_30_pro	0.1645	0.0457	3.602	< .001
Subdomain 2b: Administration & Governance	item_31	0.3645	0.0488	7.475	< .001
	item_32	0.6766	0.0584	11.584	< .001
	item_33	0.3474	0.0441	7.873	< .001
	item_34	0.5704	0.0524	10.88	< .001
Domain 3: Parent and Community Engagement	item_35	0.4579	0.0538	8.507	< .001
	item_36	0.3315	0.0631	5.254	< .001
	item_37	0.3946	0.0546	7.229	< .001
	item_38	0.395	0.0498	7.924	< .001

Table 14. Factor Loadings

Factor	Indicator	Estimate	SE	Z	p
	item_39	0.4385	0.0457	9.592	< .001
	item_40	0.1519	0.0351	4.33	< .001
	item_41	0.3268	0.0415	7.871	< .001
	item_42	0.2532	0.0465	5.442	< .001
	item_43	0.0483	0.0233	2.073	0.038
Domain 5: Nutrition & Safety	item_50	0.1445	0.0531	2.719	0.007
	item_51	0.2327	0.0518	4.494	< .001
	item_52	0.3038	0.0524	5.8	< .001
	item_53	0.2734	0.0532	5.141	< .001

Of note is that a few factors displayed high covariances (see Table 15), indicating possible multicollinearity issues. Subdomains 1a and 1b, 1b and 1c, 1b and 1d, and 2a and 2a had high covariance values. However, this is expected as these subdomains belong to the same overarching Domain. However, Subdomain 2b (Administration & Governance) covaried highly with Domains 3 (Parent & Community Engagement) and 5 (Nutrition & Safety) indicating a possible lack of discriminant validity. This is not considered a major risk in the short run but it is something worth reviewing in future. We believe that these covariances should not detract from the internal validity of the measure.

Table 15. Factor covariances.

	Sub. 1a	Sub. 1b	Sub. 1c	Sub.1d	Sub. 2a	Sub. 2b	Domain 3	Domain 5
Subdomain 1a: Equipment	1.000							
Subdomain 1b: Daily Program	0.830	1.000						
Subdomain 1c: Staff-Child Interaction	0.630	0.803	1.000					
Subdomain 1d: Planning & Assessment	0.646	0.708	0.549	1.000				
Subdomain 2a: Staffing	0.675	0.646	0.441	0.734	1.000			
Subdomain 2b: Administration & Governance	0.596	0.618	0.381	0.480	0.850	1.000		
Domain 3: Parent & Community Engagement	0.604	0.593	0.360	0.600	0.828	0.744	1.000	
Domain 5: Nutrition & Safety	0.563	0.623	0.302	0.649	0.897	0.733	1.153	1.000

Section E | Reliability

Cronbach's alpha (CA) is a measure of internal consistency, determining whether items on a test are measuring the same characteristic. Technically it may not make sense to do Cronbach alpha because not all responses on the QASS are single source. However, if we treat all the indicators as single source - the CA could be informative. First, we analysed reliability at the subdomain/domain level, and then the scale as a whole.

Table 16 below represents the accepted interpretation of Cronbach alpha values.

Table 16. Cronbach alpha interpretation

Interpretation	Cronbach's alpha
Excellent	> 0.9
Good	> 0.8
Acceptable	> 0.7
Questionable	> 0.6
Poor	> 0.5
Unacceptable	< 0.5

A Cronbach's **alpha of > 0.7 is considered acceptable reliability**. Cronbach alpha values at the subdomain and domain level are presented in Table 17. At the subdomain levels, Sub-domain 1b and 2b demonstrated Cronbach alpha levels below 0.7, indicating questionable reliability. Subdomains 1a and 2a have alpha values < 0.6, indicating poor reliability. However, it should be noted as a domain overall, Domains 1 (Cronbach alpha = 0.870) and 2 (Cronbach alpha = 0.705) show acceptable to good reliability. Domain 3 also shows questionable reliability with an alpha of 0.651. However, Domains 4 and 5 show unacceptable reliability – this could be due to the fact that these domains only comprise few items and lack a variety of variance (and a significant proportion of items making up these domains scored as zero). Correlation heatmaps can be found in Appendix D.

Table 17. QASS Reliability

	Cronbach's alpha
Domain 1: Learning Programme	0.870
Sub-domain 1a: Equipment; Play Materials & Layout	0.571
Sub-domain 1b: Daily Programme	0.698
Sub-domain 1c: Staff-Child Interaction	0.789
Sub-domain 1d: Planning & Assessment	0.716
Domain 2: Leadership & Management	0.705
Sub-domain 2a: Staffing	0.525
Sub-domain 2b: Administration & Governance	0.663
Domain 3: Parent & Community Engagement	0.651
Domain 4: Inclusiveness	0.327
Domain 5: Nutrition & Safety	0.327

Table 18 below shows how Cronbach alpha values change with removal of items from the QASS. Values highlighted in red are items, which if removed, would increase the alpha value of their respective subdomain/domains. However, these values should be interpreted with caution because in most cases deleting an item may have an impact on the content validity of a domain while not necessarily leading to a large improvement in overall CA. Two exceptions may be Item 2 in domain 1a, and item 48 in domain 4.

Table 18. Change in Cronbach Alpha by deleting items

	Cronbach's α if item dropped		Cronbach's α if item dropped
Subdomain 1a: Equipment; Play Materials & Layout	Cronbach's α for subdomain = 0.571	Subdomain 2a: Staffing	Cronbach's α for subdomain = 0.525
item_1	0.476	item_23	0.483
item_2	0.626	item_24	0.514
item_3	0.392	item_25	0.448
item_4	0.503	item_26_pro	0.562
Subdomain 1b: Daily Programme	Cronbach's α for subdomain = 0.693	item_27_pro	0.449
item_5	0.659	item_28_pro	0.432
item_6	0.643	item_29	0.476
item_7	0.639	item_30_pro	0.544
item_8	0.699	Subdomain 2b: Administration & Governance	Cronbach's α for subdomain = 0.663
item_9	0.670	item_31	0.658
item_10	0.636	item_32	0.543
Subdomain 1c: Staff-Child Interaction	Cronbach's α for subdomain = 0.789	item_33	0.620
item_11	0.761	item_34	0.538
item_12	0.750	Domain 3: Parent & Community Engagement	Cronbach's α for domain = 0.651
item_13_pro	0.783	item_35	0.596
item_14	0.816	item_36	0.624
item_15	0.754	item_37	0.607
item_16	0.750	item_38	0.613
item_17	0.754	item_39	0.598
item_18	0.754	item_40	0.650
		item_41	0.616
		item_42	0.636
Subdomain 1d: Planning & Assessment	Cronbach's α for subdomain = 0.716	item_43	0.655
item_19	0.681	Domain 4: Inclusiveness	Cronbach's α for domain = 0.327
item_20	0.617	item_44	0.301
item_21	0.634	item_45_pro	0.249
item_22	0.679	item_46	0.242
		item_47	0.194
		item_48_pro	0.405
		item_49	0.301
		Domain 5: Nutrition & Safety	Cronbach's α for domain = 0.327
		item_50	0.272
		item_51	0.195
		item_52	0.314
		item_53	0.287

Item-total correlations (see Table 19) were computed as they measure how uniform items in a test are in measuring a single concept and how well an item differentiates between participants. Ideally, we want significant and high item-total correlations ($r > 0.4$; Briggs & Cheek, 1986).

All item-total correlations were statistically significant, except for items 26 and 48 (which were shown earlier to be problematic). Furthermore, several items had correlations less than 0.4 (highlighted in orange in Table 19) and could be considered problematic. Although many items reported item total correlations lower than 0.40, only a few items had very low values. In particular item 26 and 48 reported relatively low item total correlations. These items were flagged in previous analyses as well.

Table 19. QASS Item-Total correlations

Item	Item-Total Correlation	Item	Item Total Correlation	Item	Item Total Correlation
item_1	0.496	item_19	0.564	item_37	0.367
item_2	0.35	item_20	0.515	item_38	0.479
item_3	0.532	item_21	0.534	item_39	0.472
item_4	0.466	item_22	0.462	item_40	0.252
item_5	0.508	item_23	0.36	item_41	0.403
item_6	0.524	item_24	0.172	item_42	0.315
item_7	0.49	item_25	0.519	item_43	0.137
item_8	0.378	item_26_pro	0.058	item_44	0.42
item_9	0.488	item_27_pro	0.397	item_45_pro	0.266
item_10	0.578	item_28_pro	0.398	item_46	0.317
item_11	0.436	item_29	0.532	item_47	0.438
item_12	0.508	item_30_pro	0.223	item_48_pro	0.051
item_13_pro	0.328	item_31	0.431	item_49	0.491
item_14	0.205	item_32	0.528	item_50	0.160
item_15	0.49	item_33	0.397	item_51	0.388
item_16	0.488	item_34	0.52	item_52	0.30
item_17	0.444	item_35	0.465	item_53	0.321
item_18	0.548	item_36	0.295		

Section F | Convergent Validity

In order to assess the correlations between QASS and ECERS, Pearson's correlation coefficients were calculated across the relevant subscales on the ECERS from which QASS was mapped. Statistical significance is reported alongside these correlations, as it is important to identify whether these correlations are 'by chance' or statistically significantly different from having no correlation. It should be noted that the ECERS was developed in the USA and is considered by many as the gold standard to evaluate ELPs. We were unable to run the below correlations split by ELP fee-band (fee-band 1-3 and 4-5) because the ECERS was only assessed in fee-band 1 and 2 ELPs.

Table 20. ECERS and QASS Mapping

ECERS	QASS
Subscale 1 (Space and Furnishings)	Subscale 1A (Equipment; materials and layout (4 items))
Subscale 4 (Diversity items)	Domain 4 (Inclusiveness)
Subscale 5 (Interaction)	Subscale 1C (Staff-child interactions)
Subscale 6 (Programme Structure)	Subscale 1B (Daily Programme)
Subscale 7 (Parents and staff)	Subscale 2A (Staffing)
Subscale 7 (Parents and staff)	Domain 3 (Parent and Community Engagement)

When interpreting the Pearson's correlation coefficient between the two tools, it is important to note the size of the correlation coefficient. These can be defined in the following manner:

$0.1 < r < .3$	small correlation
$0.3 < r < .5$	medium/moderate correlation
$ r > .5$	large/strong correlation

Note: Pearson correlation values are standardised measures and are therefore not impacted by the differing scales of measurement across the ECERS (7 point scale) and QASS (4 point scale).

Correlations between the ECERS and QASS are presented in Table 21. The first set of correlations indicate a moderately strong correlation between ECERS Subscale 1 and QASS Subscale 1a (0.570). This result was not significant in the Dipstick analysis but was significant for the full sample analysis ($p < 0.01$). The second set of correlations indicate a small, but positive correlation between ECERS Subscale 4 and QASS Domain 4. This result was not significant in the Dipstick analysis nor in the full sample analysis. The third set of correlations indicate a strong correlation between ECERS Subscale 5 and QASS Subscale 1c (0.640). This result was not significant in the Dipstick analysis but was significant for the full sample analysis ($p < 0.01$). The fourth set of correlations indicate a small (in the full sample) to medium (in the Dipstick analysis) correlation between ECERS Subscale 6 and QASS Subscale 1b (0.287). This result was not significant in the Dipstick analysis but was significant for the full sample analysis ($p < 0.05$). The fifth set of correlations indicate a small (in the full sample) to medium (in the Dipstick analysis) correlation between ECERS Subscale 7 and QASS Subscale 2a. This result was not significant in the Dipstick analysis but was significant for the full sample analysis ($p < 0.05$). The sixth set of

correlations indicate a small correlation between ECERS Subscale 7 and QASS Domain 3 (0.281). This result was not significant in the Dipstick analysis but was significant for the full sample analysis ($p < 0.05$). Against the foregoing background – the QASS demonstrates strong support for convergent validity with the ECERS measures across numerous dimensions. Most of the full sample results are significant at the 0.05 level of significance, the one exception being QASS dimension 4 and ECERS dimension 4. It is important that there was a lot of missing data in dimension 4 and the results should thus be interpreted with caution.

Table 21. Correlations between ECERS and QASS

ECERS Subscale 1: Space & Furnishings	QASS Subscale 1a: Equipment; Play Materials & Layout	
	Dipstick	Full sample
	0.570	0.530***
ECERS Subscale 4: Diversity	QASS Domain 4: Inclusiveness	
	Dipstick	Full sample
	0.231	0.259
ECERS Subscale 5: Interaction	QASS Subscale 1c: Staff-Child Interaction	
	Dipstick	Full sample
	0.677	0.640***
ECERS Subscale 6: Programme Structure	QASS Subscale 1b: Daily Programme	
	Dipstick	Full sample
	0.494	0.287**
ECERS Subscale 7: Parents & Staff	QASS Subscale 2a: Staffing	
	Dipstick	Full sample
	0.306	0.280**
ECERS Subscale 7: Parents & Staff	QASS Domain 3: Parent & Community Engagement	
	Dipstick	Full sample
	0.290	0.281**

** $p < 0.05$. *** $p < 0.01$.

Section G | Criterion Validity

The underlying assumption of criterion related validity is that higher quality ELPs should lead to higher levels of school readiness as measured by the ELOM measure. To this end we used the **QASS dimensions to predict the ELOM domain and total scores.**

A note on statistical significance

It will become apparent throughout this section that we pay special attention to findings that display statistical significance. In regression analysis, statistical significance plays a crucial role in helping us understand the influence of different variables (covariates) on our outcome of interest (in this case, ELOM score). The goal is often to identify which variables have a meaningful impact on the outcome. If a covariate's coefficient is statistically significant, we can be more confident that changes in that covariate are associated with changes in the outcome, and that this is not simply due to chance. **Without statistical significance, an outcome we observe might not be indicative of any underlying systematic effect or relationship,** but rather something that happened as a result of randomness or variability inherent in the data.

QASS Main Effects

Our criterion analysis begins by examining QASS as a predictor of ELOM. The standardised Beta coefficients in the regression model should indicate which independent variables are valid predictors of the ELOM total score. It is important to note that we did not have a priori hypotheses of which QASS dimensions should be predictive of ELOM domains (see Table 22 for domains). Thus, the analysis strategy regressed the ELOM total score on each of the QASS domains. This is considered the baseline model.

Table 22. ELOM and QASS Domains

ELOM		QASS	
Domain 1	Gross Motor Development	Domain 1	Learning Programme
Domain 2	Fine Motor Coordination and Visual Motor Integration	Domain 2	Programme Management and Leadership
Domain 3	Emergent Numeracy and Mathematics	Domain 3	Parent and Community Engagement
Domain 4	Cognition and Executive Functioning	Domain 4	Inclusiveness
Domain 5	Emergent Literacy and Language	Domain 5	Health and Safety

Linear regression analyses (see Table 23) were significant in predicting ELOM Total Score, Domain 2 (Fine Motor Coordination & Visual Motor Integration), and Domain 4 (Cognition & Executive Functioning). In all 3 of these models, the only **significant predictor was QASS Domain 2 (Leadership & Management)** (see Table 24).

Table 23. Model summary of regression fit across 6 models

Model	R	R ²	Adjusted R ²	Overall Model Test			
				F	df1	df2	p
Model 1 - ELOM Total Score	0.222	0.0492	0.0349	3.44	4	266	0.009
Model 2 - Domain 1: Gross Motor Development	0.146	0.0214	0.00670	1.46	4	266	0.216

Model 3 - Domain 2: Fine Motor Coordination & Visual Motor Integration	0.304	0.0922	0.0786	6.76	4	266	<.001
Model 4 - Domain 3: Emergent Numeracy & Mathematics	0.187	0.0348	0.0203	2.40	4	266	0.050
Model 5 - Domain 4: Cognition & Executive Functioning	0.279	0.0779	0.0640	5.62	4	266	<.001
Model 6 - Domain 5: Emergent Literacy & Language	0.142	0.0201	0.00541	1.37	4	266	0.246

Results from Table 23 suggest that all the regression models were significant except model 2 (ELOM Gross Motor Development Dependent Variable) and Model 6 (ELOM Emergent Literacy and Language). All the other regression models were statistically significant – that is to say that **the QASS dimension(s) scores were predictive of ELOM domain score(s)**. Table 24 below provides more information regarding the specific QASS dimensions that were predictive of ELOM dimensions.

Table 24. QASS Predictors of ELOM scores

			95% Confidence Interval					95% Confidence Interval	
Predictor	Estimate	SE	Lower	Upper	t	p	Stand. Estimate	Lower	Upper
Model 1: ELOM Total Score									
Intercept	35.81	3.55	28.82	42.81	10.08	<.001			
D1: Learning Programme	-0.01	0.08	-0.17	0.14	-0.16	0.88	-0.01	-0.15	0.13
D2: Leadership & Management	0.48	0.19	0.10	0.86	2.50	0.01	0.19	0.04	0.34
D3: Parent & Community Engagement	0.25	0.27	-0.28	0.78	0.92	0.36	0.07	-0.08	0.22
D5: Nutrition & Safety	-0.12	0.47	-1.05	0.81	-0.25	0.80	-0.02	-0.15	0.12
Model 2: ELOM Domain 1 (Gross Motor Development)									
Intercept	6.58	1.02	4.57	8.59	6.45	<.001			
D1: Learning Programme	-0.01	0.02	-0.06	0.03	-0.49	0.63	-0.03	-0.18	0.11
D2: Leadership & Management	0.05	0.06	-0.06	0.16	0.92	0.36	0.07	-0.08	0.22
D3: Parent & Community Engagement	0.04	0.08	-0.11	0.19	0.51	0.61	0.04	-0.11	0.19
D5: Nutrition & Safety	0.17	0.14	-0.10	0.44	1.25	0.21	0.09	-0.05	0.23
Model 3: ELOM Domain 2 (Fine Motor Coordination & Visual Motor Integration)									
Intercept	7.41	0.77	5.90	8.92	9.67	<.001			
D1: Learning Programme	0.01	0.02	-0.03	0.04	0.32	0.75	0.02	-0.11	0.16

Table 24. QASS Predictors of ELOM scores

Predictor	Estimate	SE	95% Confidence Interval		t	p	Stand. Estimate	95% Confidence Interval	
			Lower	Upper				Lower	Upper
D2: Leadership & Management	0.14	0.04	0.06	0.22	3.40	<.001	0.25	0.11	0.40
D3: Parent & Community Engagement	0.06	0.06	-0.05	0.18	1.09	0.28	0.08	-0.07	0.23
D5: Nutrition & Safety	-0.05	0.10	-0.25	0.15	-0.47	0.64	-0.03	-0.17	0.10
Model 4: ELOM Domain 3 (Emergent Numeracy & Mathematics)									
Intercept	6.51	0.84	4.85	8.16	7.76	<.001			
D1: Learning Programme	0.00	0.02	-0.03	0.04	0.26	0.80	0.02	-0.12	0.16
D2: Leadership & Management	0.08	0.05	-0.01	0.17	1.80	0.07	0.14	-0.01	0.29
D3: Parent & Community Engagement	0.05	0.06	-0.07	0.18	0.82	0.42	0.06	-0.09	0.22
D5: Nutrition & Safety	-0.01	0.11	-0.23	0.21	-0.11	0.92	-0.01	-0.15	0.13
Model 5: ELOM Domain 4 (Cognition & Executive Functioning)									
Intercept	3.81	0.84	2.16	5.46	4.55	<.001			
D1: Learning Programme	0.00	0.02	-0.04	0.03	-0.12	0.90	-0.01	-0.15	0.13
D2: Leadership & Management	0.13	0.05	0.04	0.22	2.93	<.001	0.22	0.07	0.37
D3: Parent & Community Engagement	0.08	0.06	-0.05	0.20	1.21	0.23	0.09	-0.06	0.24
D5: Nutrition & Safety	0.02	0.11	-0.20	0.24	0.16	0.87	0.01	-0.12	0.15
Model 6: ELOM Domain 5 (Emergent Literacy & Language)									
Intercept	11.51	0.94	9.65	13.36	12.23	<.001			
D1: Learning Programme	-0.01	0.02	-0.05	0.03	-0.45	0.65	-0.03	-0.17	0.11
D2: Leadership & Management	0.07	0.05	-0.03	0.17	1.44	0.15	0.11	-0.04	0.26
D3: Parent & Community Engagement	0.02	0.07	-0.13	0.16	0.22	0.83	0.02	-0.14	0.17
D5: Nutrition & Safety	-0.25	0.13	-0.49	0.00	-1.99	0.05	-0.14	-0.28	0.00

In Model 1, the ELOM total score was regressed on the QASS dimensions. The only significant predictor of the ELOM total score was **Dimension 2 of QASS (Leadership and Management, $\beta = 0.19$, $p < 0.01$)**. All the **other predictors were not statistically significant**.

In Model 2, ELOM domain 1 (Gross Motor Development) was regressed on the QASS dimensions. None of the individual predictors were statistically significant.

In Model 3, ELOM domain 2 (Fine Motor Coordination and Visual Motor Integration) was regressed on the QASS dimensions. Once again, the only significant predictor was Dimension 2 of QASS (Leadership and Management, $\beta = 0.25$, $p < 0.01$). The same result was found for Model 5 (Domain 4 Cognitive and Executive Functioning as dependent variable). No single predictor was effective in predicting Domain 5 of ELOM (Emergent Literacy and Language), except Dimension 5 (Nutrition & Safety).

These results suggest some predictive validity for the QASS where Leadership and Management domain seem to be predictive of a number of ELOM domains including ELOM Total Score; Fine Motor Coordination and Visual Motor Integration; Emergent Numeracy; Cognition and Executive Functioning.

Multiple regression analysis

It should be noted that the regression models presented in the prior subsection are 'fundamental' regressions looking only at the main effect of QASS on ELOM score. These models generally have low R^2 values, meaning that a lot of the variance in ELOM score has not been explained (i.e. there are other factors that have not been controlled for that are also influencing ELOM score). In order to address this, we proceed with more comprehensive regression analyses that control for fee band and province⁹. These comprehensive analyses show a substantial improvement in the R^2 values, while the take-home message regarding the relationship between QASS and ELOM remains the same: **Domain 2 is the key significant predictor.**

In Table 25, we see the model summary statistics for the regression including fee band and province as covariates, which indicate a substantial improvement in terms of the variance explained. Models discussed in the previous subsection had R^2 values between 2 and 10 percent (meaning that 2-10% of the variance in ELOM score is being explained by the model). With the inclusion of province and fee band as covariates, we see the R^2 value increases to a more acceptable 18.59% and an adjusted R^2 of 15.46%. All this being said, there are definitely other factors influencing the variance of ELOM score (as there is a remaining ~80% of variance that is left unexplained) which have not been controlled for and provide an opportunity for further research.

When Fee Band and Province were added as covariates (see Table 26), QASS Leadership and Management (Estimate = 0.365, $p = 0.0465$), Fee Band and Province were significant predictors of ELOM Total Score. This means that, when holding factors such as fee band and province constant, we are still able to observe a significant (at the 5% level) relationship between QASS Domain 2 and ELOM Total score - where a 1 point increase in the QASS Domain 2 score is associated with a ~0.37 point increase in ELOM score. Relative to fee band 1 (the reference category), fee band 5 was significantly associated with a ~10 point higher ELOM score¹⁰. Interestingly, province was a highly significant (at the 1% level) covariate. Relative to Kwa-Zulu Natal ELPs ($n=178$), Mpumalanga ELPs ($n=42$) were associated with lower ELOM scores while Western Cape ELPs (68) were associated with higher ELOM scores¹¹. Results for the Western Cape in comparison to Kwa-Zulu Natal are reflective of the nationally representative findings seen in the *Thrive by five 2021*, while results for Mpumalanga tend to differ.

Table 25. Model summary of regression fit

Overall Model Test

⁹We also explored controlling for the number of assistants, assistants per age group and geotype, however these variables did not add much explanatory power to the model and were not significant predictors.

¹⁰Note: Fee band 5 has a limited sample size of 9 ELPs and this result should be interpreted tentatively until further research by fee-band is conducted.

¹¹Another important caveat is that this validation study was not designed to be geographically representative and the sample size across provinces are not proportional to size - therefore provincial relationships with ELOM score are only reflective of the ELPs involved in this study and should not be generalised.

Model	R ²	Adjusted R ²	F	df1	df2	p
1	0.1859	0.1546	5.94	10	260	<.001

Table 26. QASS Predictors of ELOM scores controlling for Fee Band (reference category: Fee Band 1) and Province (reference category: Kwa-Zulu Natal)

Predictor	Estimate	SE	95% Confidence Interval		t	p
			Lower	Upper		
D1: Learning Programme	-0.0652	-0.0775	-0.218	0.0874	-0.841	0.401
D2: Leadership & Management	0.365**	-0.182	0.0057	0.724	2.000	0.0465
D3: Parent & Community Engagement	0.41	-0.268	-0.117	0.937	1.531	0.127
D5: Nutrition & Safety	-0.295	-0.446	-1.173	0.584	-0.661	0.509
Fee band 2	-1.735	-1.609	-4.903	1.433	-1.078	0.282
Fee band 3	-0.257	-2.534	-5.248	4.733	-0.102	0.919
Fee band 4	1.016	-3.149	-5.184	7.216	0.323	0.747
Fee band 5	10.72**	-4.167	2.512	18.92	2.572	0.0107
Mpumalanga	-5.031**	-2.127	-9.219	-0.843	-2.366	0.0187
Western Cape	6.226***	-2.115	2.062	10.39	2.944	0.0035
Intercept	37.82***	-3.429	31.07	44.57	11.03	0

Further Analysis of Domain 1 as a predictor of ELOM

We had expected that QASS Domain 1 (Learning Programme) would be a significant predictor of ELOM score, however when controlling for the other domains as well as the additional covariates, we did not observe any significant relationship. We therefore conducted further ad-hoc analyses where the ELOM total score was regressed on the QASS domain 1 and subdomain scores to isolate possible significant relationships¹².

We begin by exploring simple correlations between domain 1 and subdomain 1a, 1b, 1c and 1d with ELOM total score. Apart from Subdomain 1c, each dimension from domain 1 significantly correlates with ELOM Score when assessed in isolation, with Subdomain 1b showing both the strongest correlation in terms of size and statistical significance (at the 1% level).

Table 27. QASS Domain 1 and Subdomain 1a, 1b, 1c, 1d correlation with ELOM Score

¹²These further exploratory analyses are largely for interest and did not form part of the initial validation design.

QASS Dimension	Correlation
Domain 1	0.1078*
Subdomain 1a	0.1375**
Subdomain 1b	0.1641***
Subdomain 1c	-0.0076
Subdomain 1d	0.1042*

p < 0.05. *p < 0.01.

When we include these QASS subdomains (1a, 1b, 1c and 1d) in the full regression against ELOM total Score, controlling for fee band and province, **we do not observe a significant relationship.**

Table 28. Model summary of regression fit

Overall Model Test						
Model	R	R ²	Adjusted R ²	F	df1	df2
1	0.169	0.1367	0.0572	5.27	10	260

Table 29. QASS Predictors of ELOM scores controlling for Fee Band and Province

Predictor	Estimate	SE	Lower	Upper	t	p
Subdomain 1a	0.501	-0.397	-0.281	1.283	1.262	0.208
Subdomain 1b	0.241	-0.297	-0.345	0.826	0.809	0.419
Subdomain 1c	-0.269	-0.196	-0.656	0.117	-1.373	0.171
Subdomain 1d	0.0745	-0.28	-0.476	0.625	0.266	0.79
Fee band 2	-1.373	-1.595	-4.514	1.767	-0.861	0.39
Fee band 3	-0.0459	-2.533	-5.034	4.943	-0.0181	0.986
Fee band 4	1.869	-3.13	-4.295	8.033	0.597	0.551
Fee band 5	10.66**	-4.22	2.349	18.97	2.526	0.0121
Mpumalanga	-4.298*	-2.192	-8.614	0.0176	-1.961	0.0509
Western Cape	6.192***	-2.233	1.795	10.59	2.773	0.00595
Intercept	43.64***	-2.943	37.85	49.44	14.83	0

We also explored each of the ELOM Domains (1,2,3,4,5) as dependent variables for this full regression model. Here, **the only significant (at the 5% level) findings were observed for ELOM Domain 2 (Fine Motor Coordination and Visual Motor Integration)** - reported in tables 30 and 31. Here, we observe that a one point increase in score in QASS Subdomain 1b is associated with a 0.153 point increase in ELOM Domain 2 score, significant at the 5% level. Relative to fee band 1, fee band 4 was strongly associated with a higher ELOM Domain 2 score, significant at the 1% level.

Table 30. Model summary of regression fit

Model	R	R ²	Adjusted R ²	Overall Model Test		
				F	df1	df2
1	0.149	0.1162	0.0572	4.55	10	260

Table 31. QASS Predictors of ELOM Domain 2 controlling for Fee Band and Province

Predictor	Estimate	SE	Lower	Upper	t	p
Subdomain 1a	0.115	-0.0887	-0.0599	0.289	1.294	0.197
Subdomain 1b	0.153**	-0.0664	0.0224	0.284	2.307	0.0218
Subdomain 1c	-0.0696	-0.0438	-0.156	0.0167	-1.588	0.114
Subdomain 1d	-0.0106	-0.0625	-0.134	0.112	-0.17	0.865
Fee band 2	-0.49	-0.356	-1.192	0.212	-1.375	0.17
Fee band 3	0.742	-0.566	-0.372	1.857	1.311	0.191
Fee band 4	1.881***	-0.699	0.504	3.259	2.69	0.00761
Fee band 5	1.08	-0.943	-0.777	2.936	1.145	0.253
Mpumalanga	-0.392	-0.49	-1.356	0.572	-0.8	0.424
Western Cape	0.274	-0.499	-0.709	1.256	0.549	0.584
Intercept	9.566***	-0.658	8.271	10.86	14.55	0

Overall, the criterion analysis showed that QASS Domain 2 was the core driver in the relationship between QASS and ELOM. **This is not to say that QASS Domain 1 is not correlated with ELOM outcomes, as we have explored above, but rather QASS Domain 1 is not the strongest predictor of ELOM score.** Further analysis on the relationship between QASS and ELOM is an important avenue for future research, however identifying a causal link (best addressed using longitudinal methods) between these two tools is beyond the scope of this study. This type of analysis may benefit from a sample with more variation (allowing for more predictive power), which is better achieved through representative random sampling accompanied by a larger sample size, as the variation in ELOM scores reflected in this analysis is slightly lower (SD ~12) than the nationally representative results found in *Thrive by five 2021* (SD ~14).

Conclusion

The goal of the project was to investigate the reliability and validity of the QASS measure. Numerous statistical techniques were used to investigate the reliability (Cronbach Alpha, and Cohen's Kappa), construct validity (Confirmatory Factor Analysis), convergent validity (correlations between QASS dimensions and ECERS dimensions), predictive validity (regression between ELOM dimensions and QASS dimensions).

In general, **strong support was found for the reliability of the measure** by means of Cronbach's alpha and Cohen's Kappa. Similarly **strong support was found for the internal structure** of the measure by means of Confirmatory Factor Analyses. **Domain 4 (Inclusivity) was stripped from the CFA analyses due to the large number of missing cases** on the items that made up Domain 4. Support was found for the internal structure of the tool but the fit was poor. Some of the items - such as item 26 and item 48 - showed strain and need to be considered for revision or deletion.

Good empirical support was found for the convergent validity of the tool when compared to ECERS measures. An a priori mapping exercise identified the ECERS dimensions which should correspond to the QASS dimensions. Even with a relatively small sample and independent raters – strong support was found for the convergent validity of the measure compared to ECERS.

Finally, multiple regression was used to investigate the predictive ability of QASS dimensions and total scores to predict early school readiness by means of the ELOM measure. Although the QASS domains were able to predict ELOM total score, **most of the variance was explained by Domain 2 of the QASS** (Leadership and Management). Even when controlling for fee band and province, Domain 2 remained a strong predictor.

Recommendations

Statistical recommendations

The QASS measure demonstrates strong construct, convergent and, to a lesser extent, criterion-related validity. The majority of item content resides in the Learning programme domain. The curriculum, teaching, and classroom interactions remain at the heart of any Early Learning Programme. Domain 1, however, did not correspond with the ELOM measure when controlling for other factors, with only isolated correlations between Domain 1 and ELOM showing correspondence. Admittedly, many factors impact school readiness, many of which may be outside the control of any given ELP. However, one can control for those factors if measured (e.g. SES, household income, access to more learning activities) simultaneously as the QASS. The only dimension of QASS which was predictive of ELOM total scores in the comprehensive regression analysis was Leadership and Management. This makes sense since solid leadership and mentorship lie at the heart of a well-managed ELP, with this finding also being a strong theme identified during the qualitative research component of this report (see qualitative report for further details).

Part of the reason for the lack of strength in the relationship between Domain 1 of QASS and ELOM score may be the broad domain of dimension 1 (learning programme). To simplify the internal structure, it may be worthwhile to group the items in the learning domain into three categories (e.g., Curriculum, Teaching, and Interaction). These domains can then be used in a regression model to predict learning outcomes with higher degrees of success compared to a relatively large multidimensional factor structure.

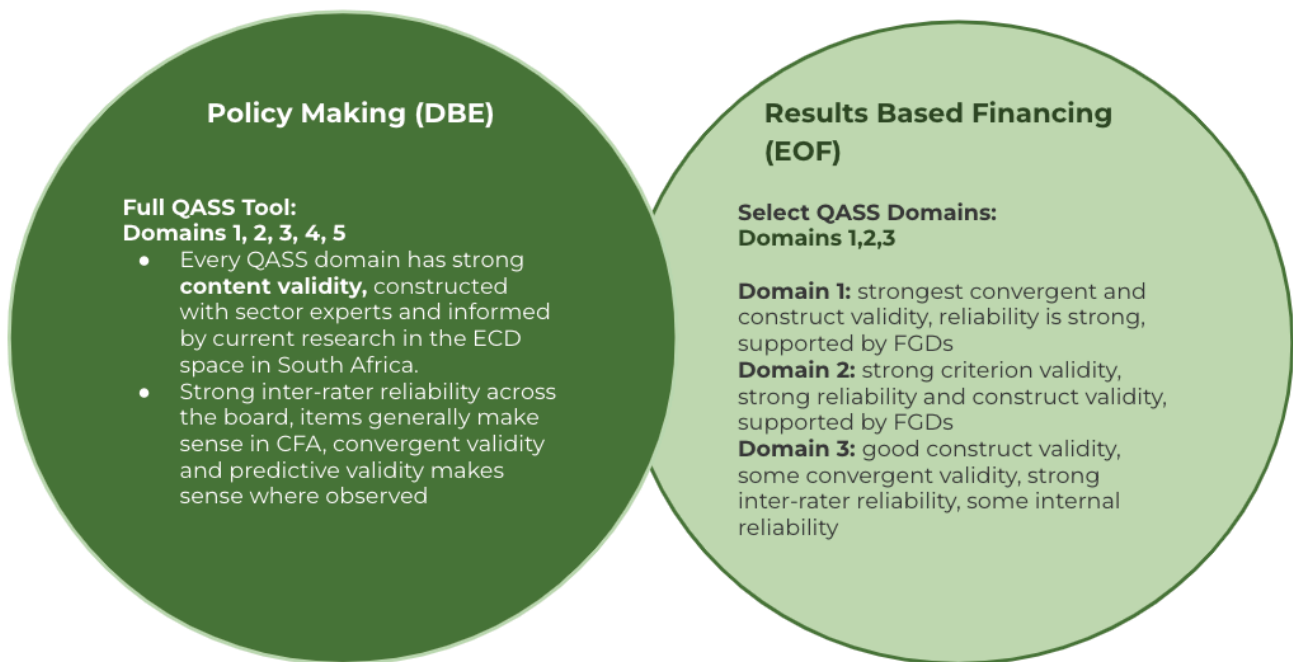
Certain items allowed for an "N/A" option, where items were pro-rated (given a 3) so as not to artificially penalise ELPs where the item was inapplicable. The many missing/inapplicable cases in Domain 4, however, meant that the pro-rated items caused the domain to have a severely restricted range and poor fit. As such, we removed the dimension from the regression and CFA analyses due to the large

number of missing data. We recommend that the items remain in the measure (since they are critically important) but they should only be completed by ELPs for which the items are applicable.

Use case recommendations

In addition to the statistical recommendations outlined in the previous subsection, it is important for us to highlight which parts of the QASS are particularly fit for purpose. We understand that there are currently two key, immediate use cases for the QASS tool: policymaking for the Department of Basic Education (DBE) and outcomes based financing work with the Education Outcomes Fund (EOF). With this in mind, our team has put together a broad conceptualisation of how the QASS may be applied in practice, based on this statistical report:

Conceptualisation of QASS usage in practice



The strong **content validity of QASS**, based on iterations of engagement with key sector stakeholders and researchers, combined with the overarching findings in this statistical report that show QASS is **overall a reliable tool that robustly correlates with other established tools such as the ECERs and ELOM**, makes the tool suitable to inform policy making needs for DBE. Results based financing work, however, requires robustness of granular metrics that may be used to trigger payments. With this in mind, we advise that further **results based financing research focuses on QASS Domains 1,2 and 3** - domains exhibiting the strongest statistical reliability and validity throughout this report.

In addition to the above, **further research and/or refinement should be conducted for Domain 4 (Inclusiveness)**, as this domain has been subject to severe sample size constraints for construct validity in this analysis (due to the many cases of inapplicability). As such, the items in Domain 4 have not necessarily had a 'fair' test for construct validity and it may be premature to remark conclusively on domain 4's applicability to use cases such as results based financing.

Appendix

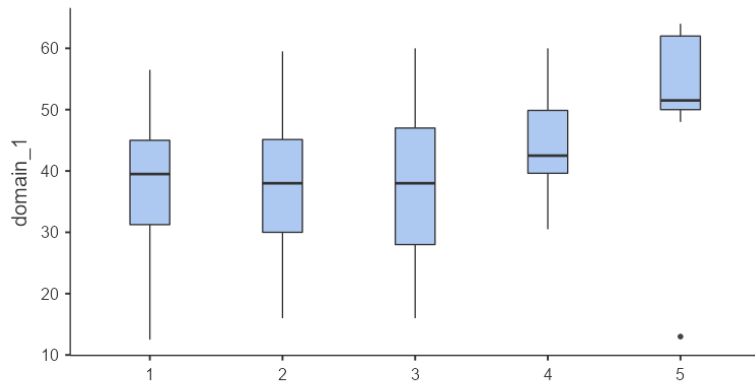
A. Demographics broken down by ELP fee-band

	Fee-band	Counts	% of Total
Geo Type			
Rural	1-3	133	46.5%
	4-5	16	5.6%
Urban	1-3	122	42.7%
	4-5	15	5.2%
ELP Type			
ECD Centre	1-3	253	87.8%
	4-5	28	9.7%
Playgroup	1-3	4	1.4%
	4-5	3	1.0%
Province			
Kwazulu-Natal	1-3	171	59.4%
	4-5	7	2.4%
Mpumalanga	1-3	35	12.2%
	4-5	7	2.4%
Western cape	1-3	51	17.7%
	4-5	17	5.9%
Registration			
None of these	1-3	7	2.4%
	4-5	3	1.0%
Registered as NPO	1-3	233	80.9%
	4-5	8	2.8%
NPO application in progress	1-3	5	1.7%
	4-5	0	0.0%
Private	1-3	12	4.2%
	4-5	20	6.9%
Bank			
No	1-3	29	10.1%
	4-5	4	1.4%
Yes	1-3	228	79.2%
	4-5	27	9.4%
Age Grouped			
No	1-3	26	9.0%
	4-5	5	1.7%
Yes	1-3	231	80.2%
	4-5	26	9.0%

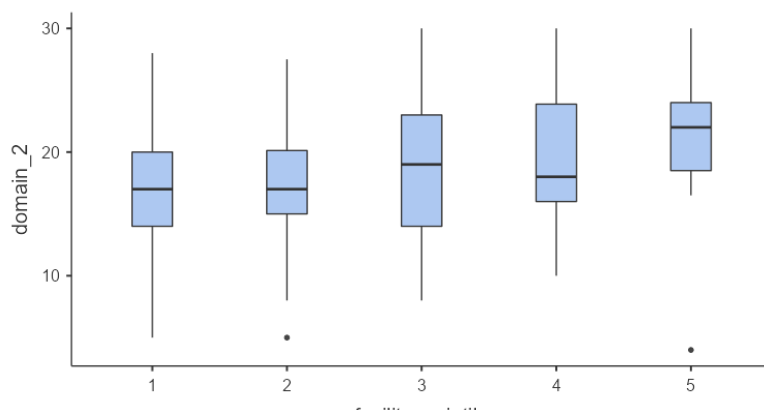
B. Boxplots of QASS, ECERS & ELOM Domain and Sub-domains by ELP fee-band

QASS

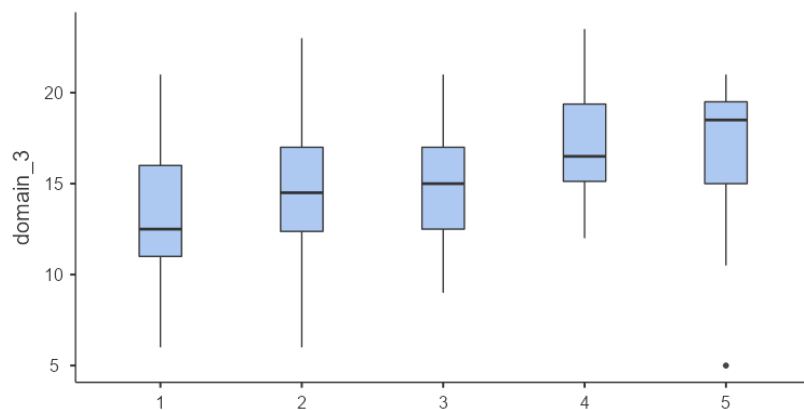
Domain 1



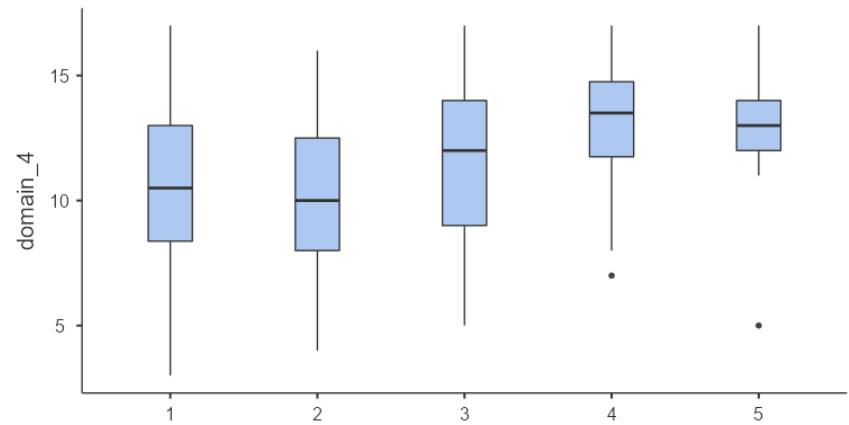
Domain 2



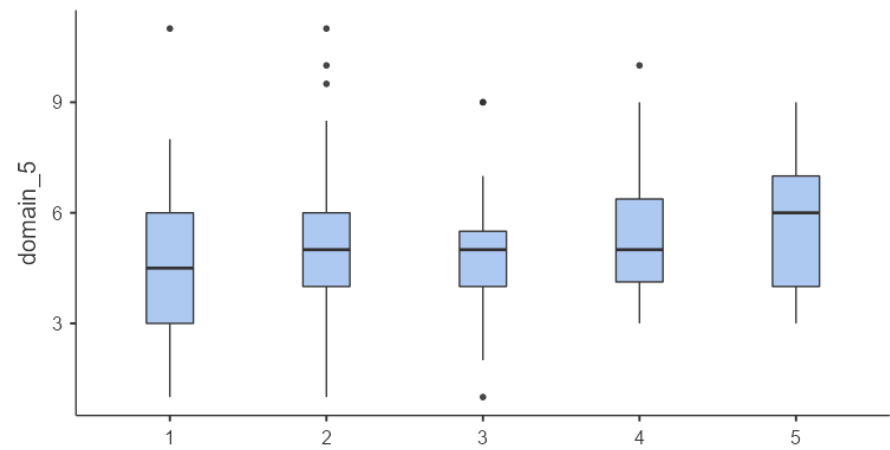
Domain 3



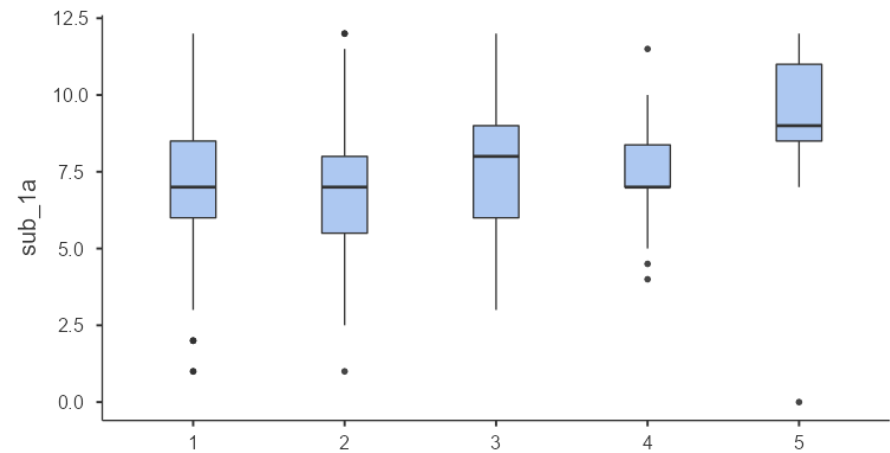
Domain 4



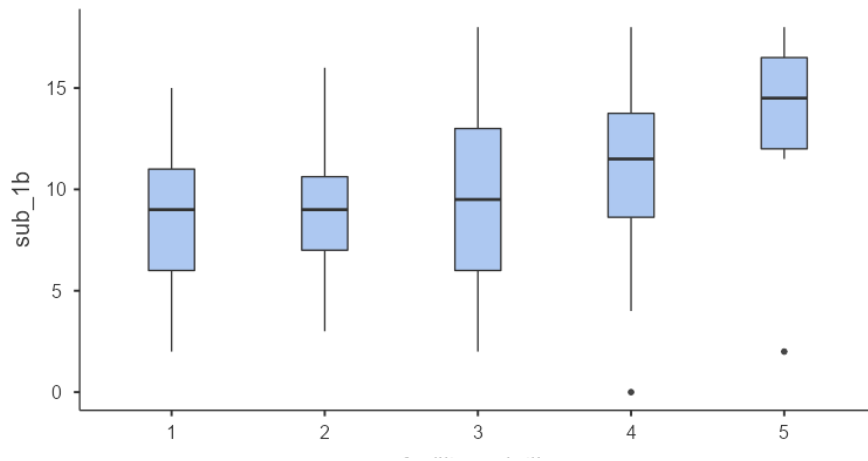
Domain 5



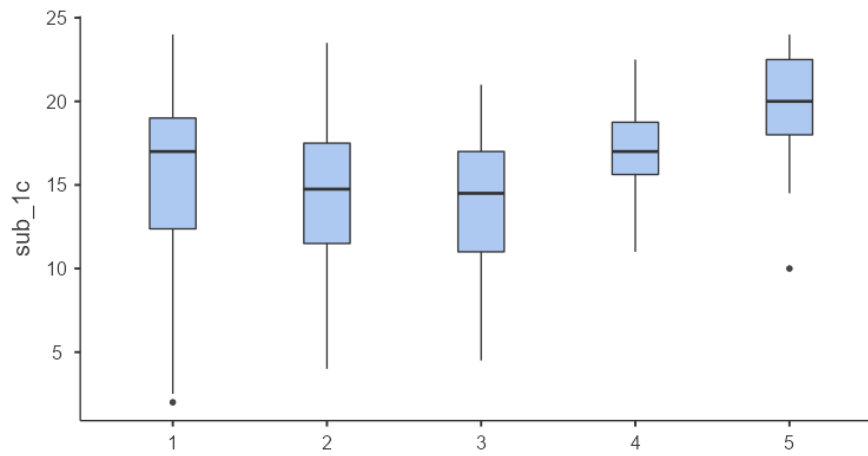
Sub-domain 1a



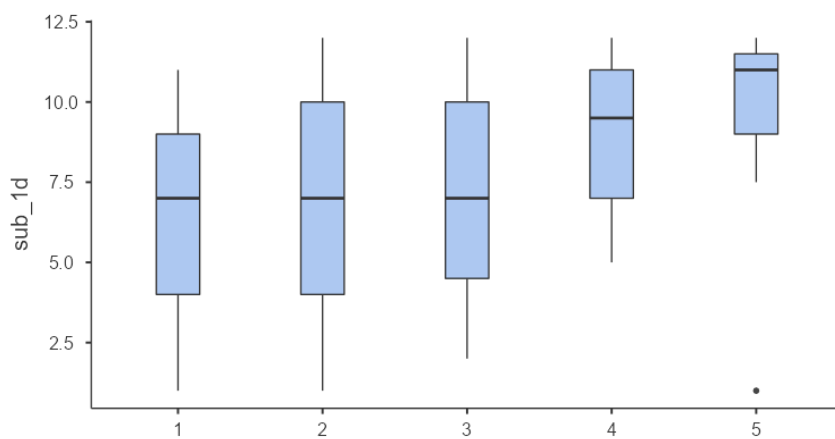
Sub-domain 1b



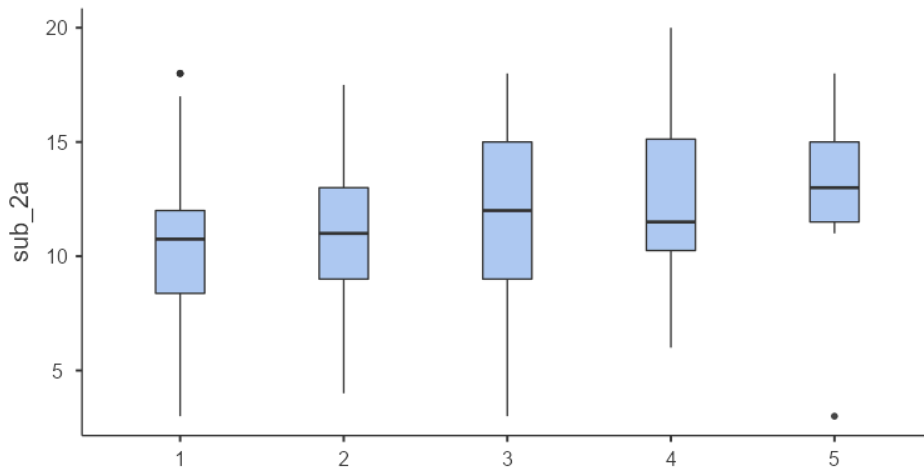
Sub-domain 1c



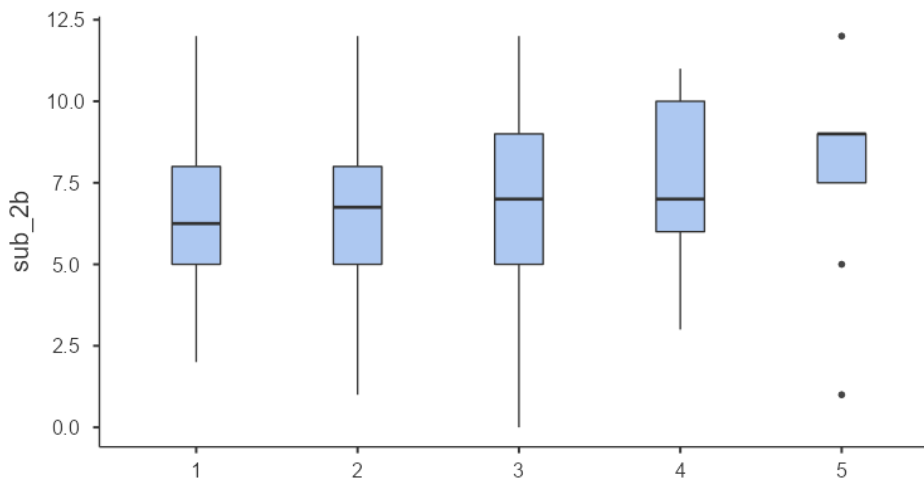
Sub-domain 1d



Sub-domain 2a

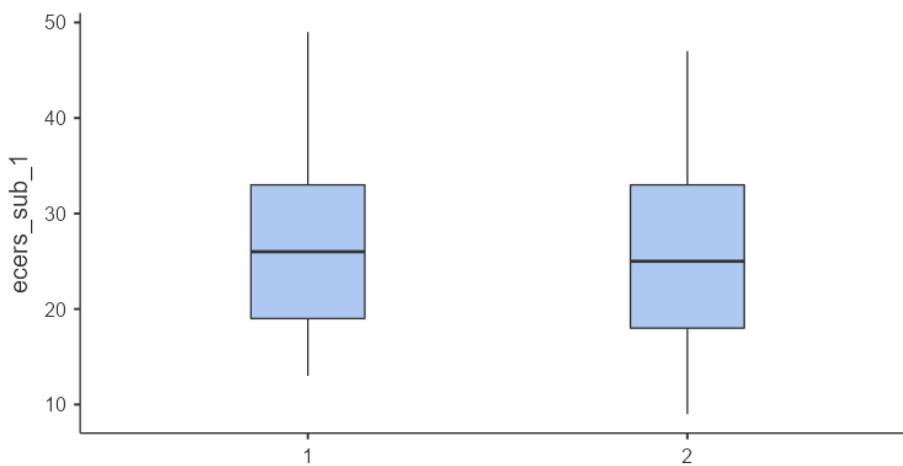


Sub-domain 2b

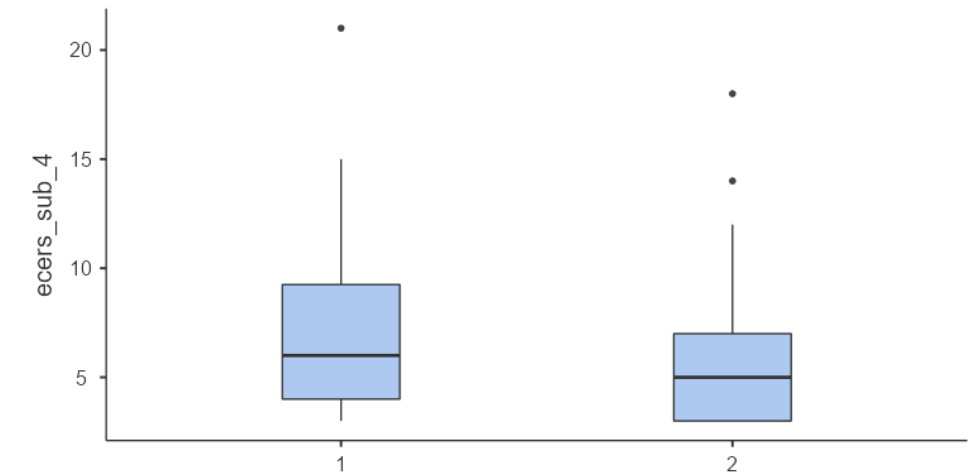


ECERS

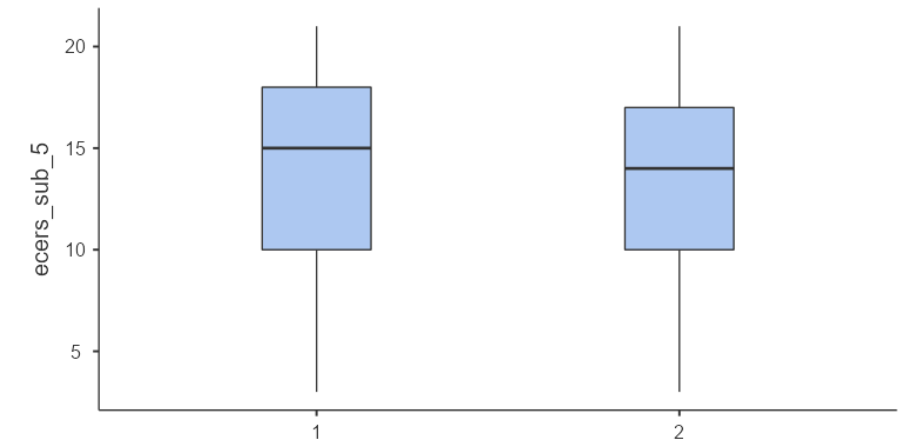
Sub-domain 1



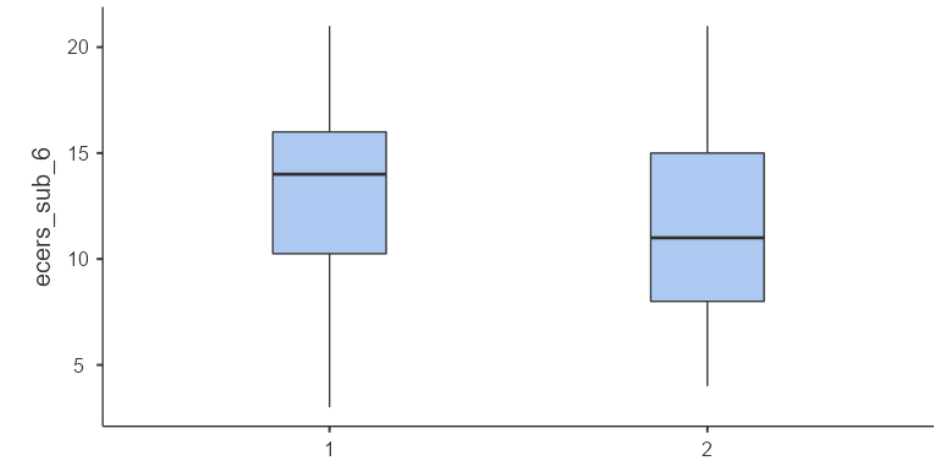
Sub-domain 4



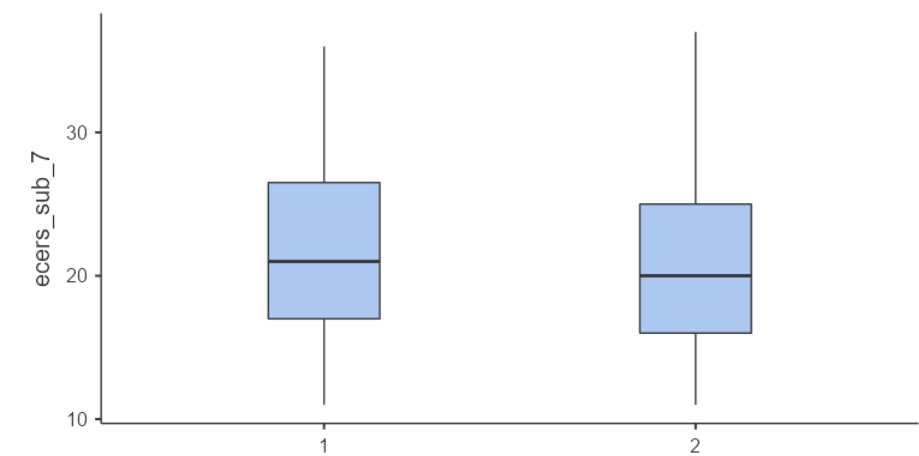
Sub-domain 5



Sub-domain 6

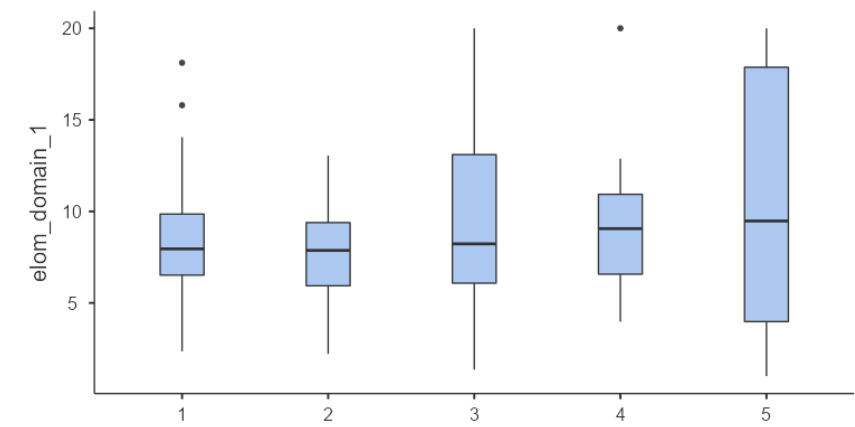


Sub-domain 7

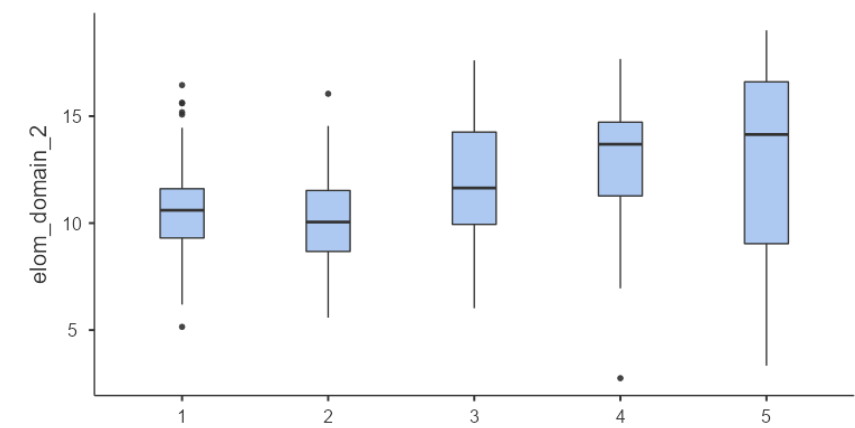


ELOM

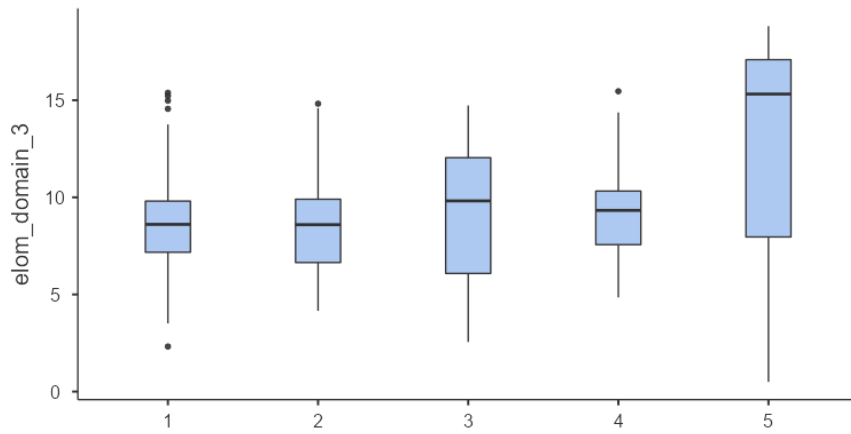
Domain 1



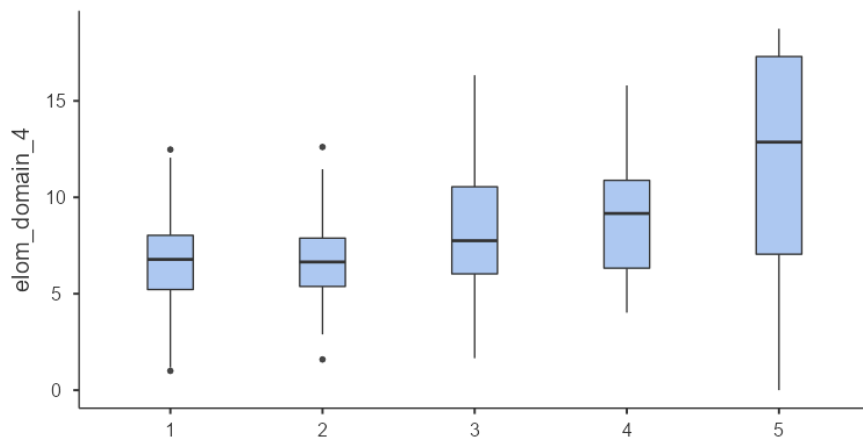
Domain 2



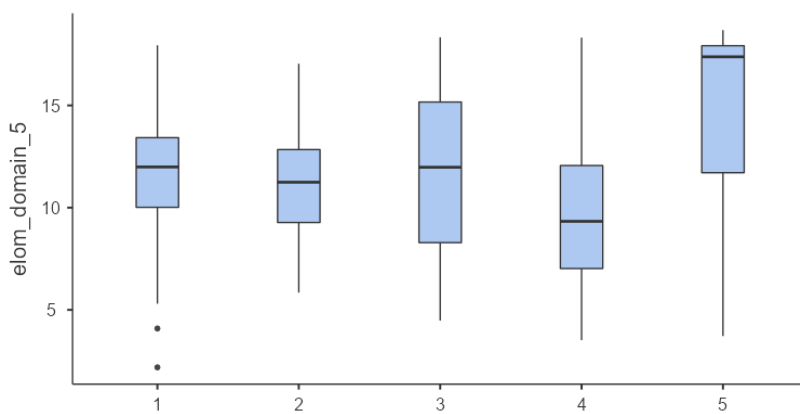
Domain 3



Domain 4



Domain 5



C. CFA at QASS Domain Level

Confirmatory Factor Analysis

Model Fit

Test for Exact Fit

χ^2	df	p
1991	1028	<.001

Fit Measures

CFI	TLI	RMSEA	RMSEA 90% CI	
			Lower	Upper
0.700	0.684	0.0570	0.0533	0.0608

Factor Loadings

Factor	Indicator	Estimate	SE	Z	p
Domain 1: Learning Programme	item_1	0.4200	0.0558	7.522	<.001
	item_2	0.4039	0.0604	6.682	<.001
	item_3	0.4056	0.0415	9.785	<.001
	item_4	0.2912	0.0391	7.441	<.001
	item_5	0.5264	0.0499	10.541	<.001
	item_6	0.5453	0.0546	9.986	<.001
	item_7	0.4217	0.0527	7.996	<.001
	item_8	0.3147	0.0575	5.469	<.001
	item_9	0.3769	0.0508	7.420	<.001
	item_10	0.4902	0.0441	11.120	<.001
	item_11	0.4397	0.0474	9.281	<.001
	item_12	0.5259	0.0478	11.013	<.001
	item_13_pro	0.4955	0.0675	7.338	<.001
	item_14	0.2142	0.0658	3.253	0.001

Factor Loadings

Factor	Indicator	Estimate	SE	Z	p
Domain 2: Leadership & Management	item_15	0.4468	0.0397	11.260	<.001
	item_16	0.6955	0.0638	10.904	<.001
	item_17	0.4943	0.0478	10.351	<.001
	item_18	0.4637	0.0387	11.976	<.001
	item_19	0.6385	0.0633	10.079	<.001
	item_20	0.4982	0.0629	7.925	<.001
	item_21	0.4636	0.0589	7.871	<.001
	item_22	0.3679	0.0582	6.318	<.001
	item_23	0.3661	0.0640	5.716	<.001
	item_24	0.1488	0.0551	2.698	0.007
	item_25	0.5106	0.0500	10.210	<.001
	item_26_pro	0.0196	0.0545	0.359	0.720
	item_27_pro	0.3151	0.0498	6.328	<.001
	item_28_pro	0.3828	0.0609	6.287	<.001
	item_29	0.3391	0.0337	10.066	<.001
	item_30_pro	0.1523	0.0455	3.345	<.001
	item_31	0.3455	0.0473	7.298	<.001
	item_32	0.6268	0.0571	10.981	<.001
	item_33	0.3202	0.0430	7.446	<.001
	item_34	0.5439	0.0503	10.804	<.001
Domain3: Parent & Community Engagement	item_35	0.4508	0.0540	8.350	<.001
	item_36	0.3327	0.0631	5.274	<.001
	item_37	0.3822	0.0547	6.988	<.001
	item_38	0.3885	0.0499	7.794	<.001

Factor Loadings

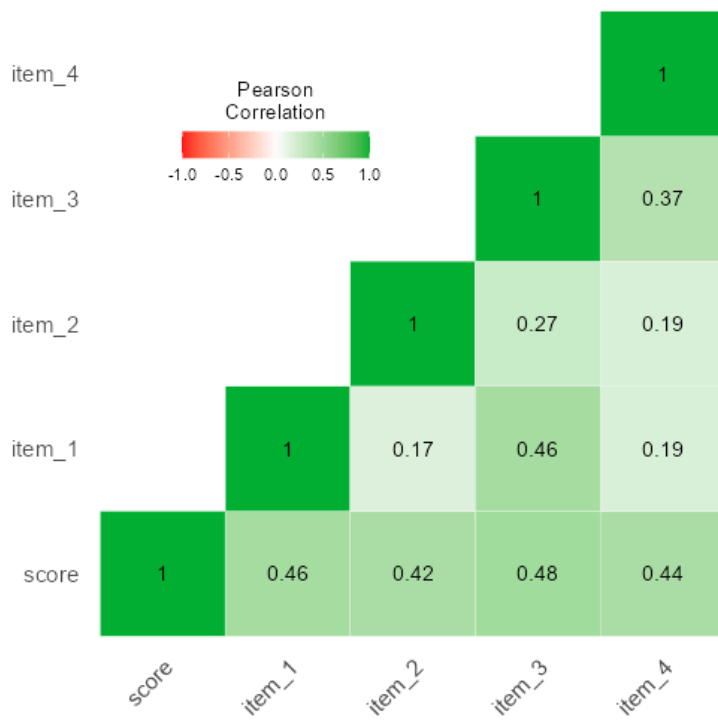
Factor	Indicator	Estimate	SE	Z	p
Domain 5: Nutrition & Safety	item_39	0.4454	0.0454	9.807	<.001
	item_40	0.1570	0.0350	4.486	<.001
	item_41	0.3269	0.0416	7.860	<.001
	item_42	0.2640	0.0460	5.733	<.001
	item_43	0.0465	0.0233	1.995	0.046
	item_50	0.1330	0.0532	2.499	0.012
	item_51	0.2296	0.0517	4.438	<.001
	item_52	0.3095	0.0526	5.887	<.001
	item_53	0.2692	0.0532	5.056	<.001

Factor Covariances QASS

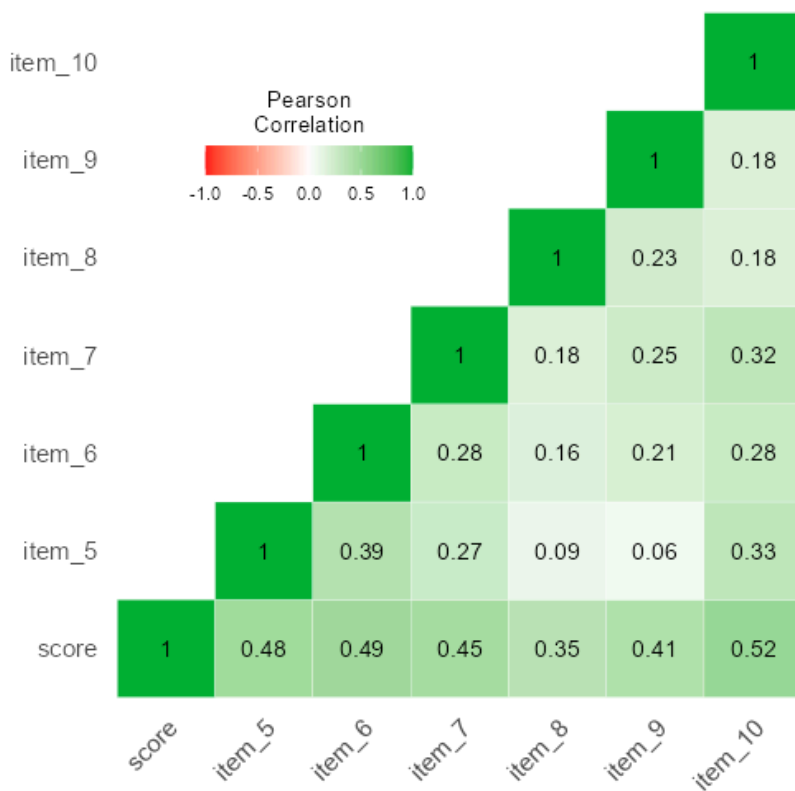
	Domain 1	Domain 2	Domain 3	Domain 5
Domain 1: Learning Programme	1.000			
Domain 2: Leadership & Management	0.610	1.00		
Domain3: Parent & Community Engagement	0.547	0.823	1.000	
Domain 5: Nutrition & Safety	0.531	0.849	1.164	1.000

D. Correlation Heatmaps

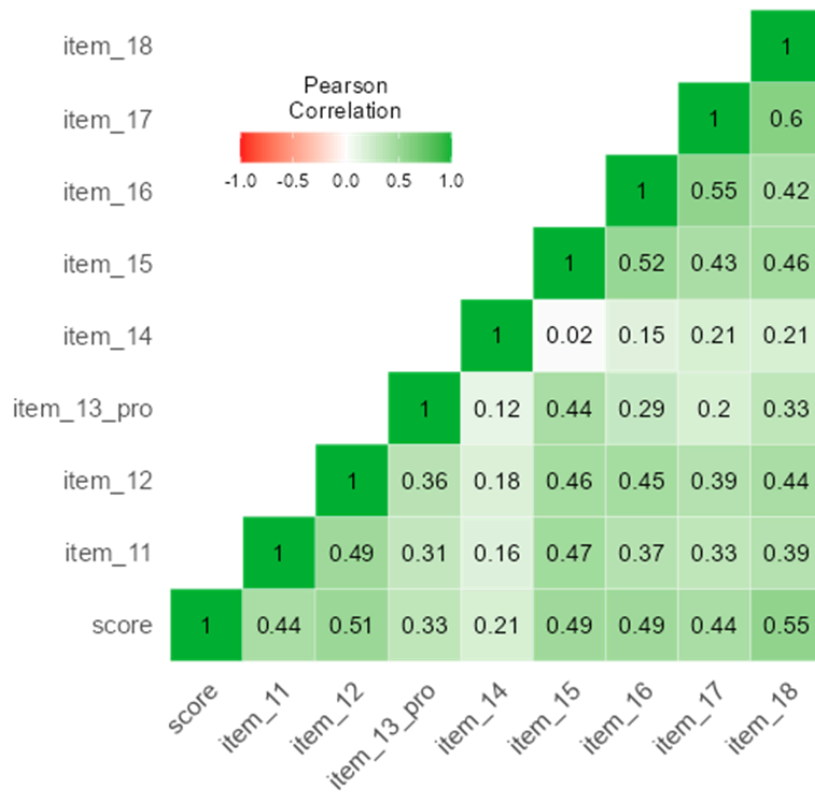
QASS Subdomain 1a



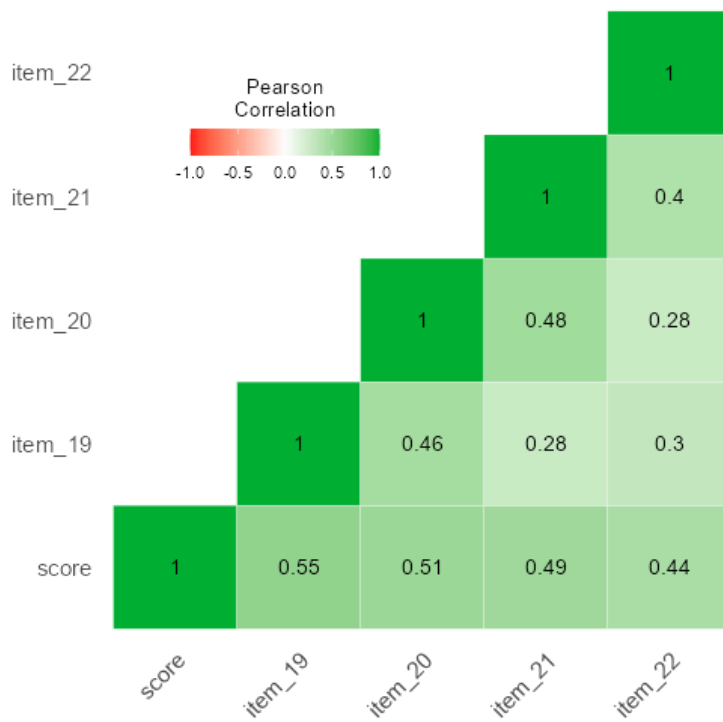
QASS Subdomain 1b



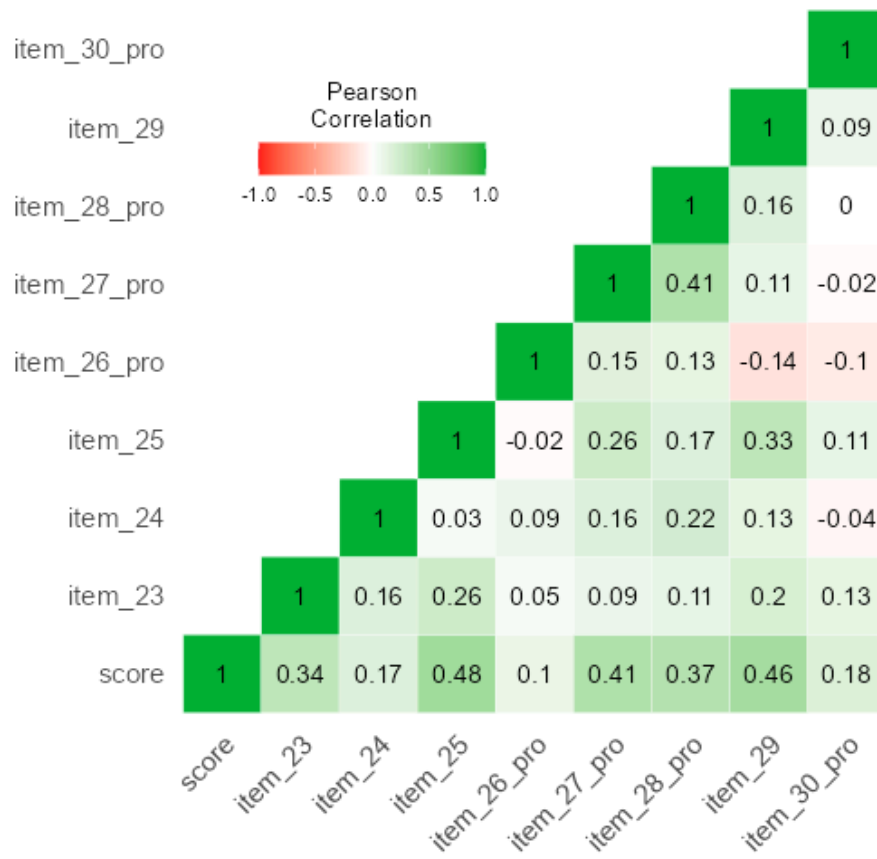
QASS Subdomain 1c



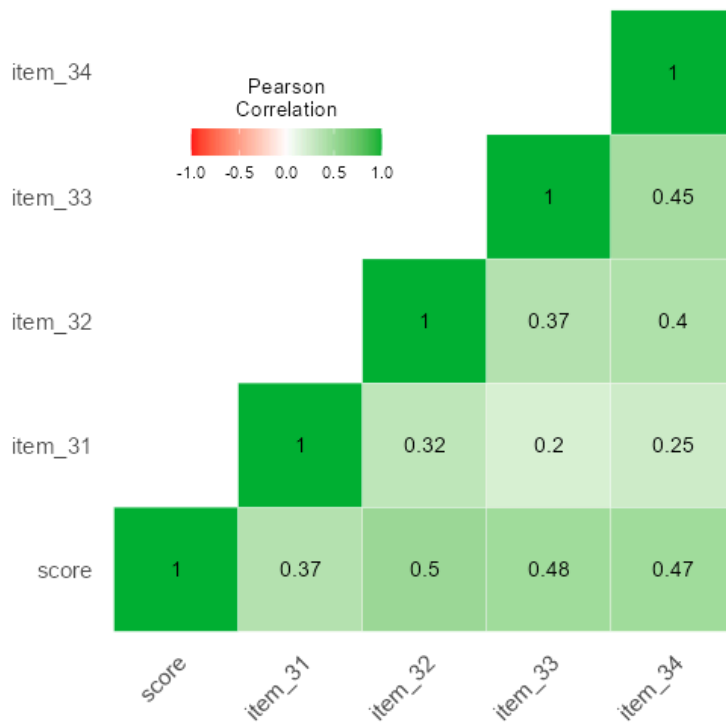
QASS Subdomain 1d



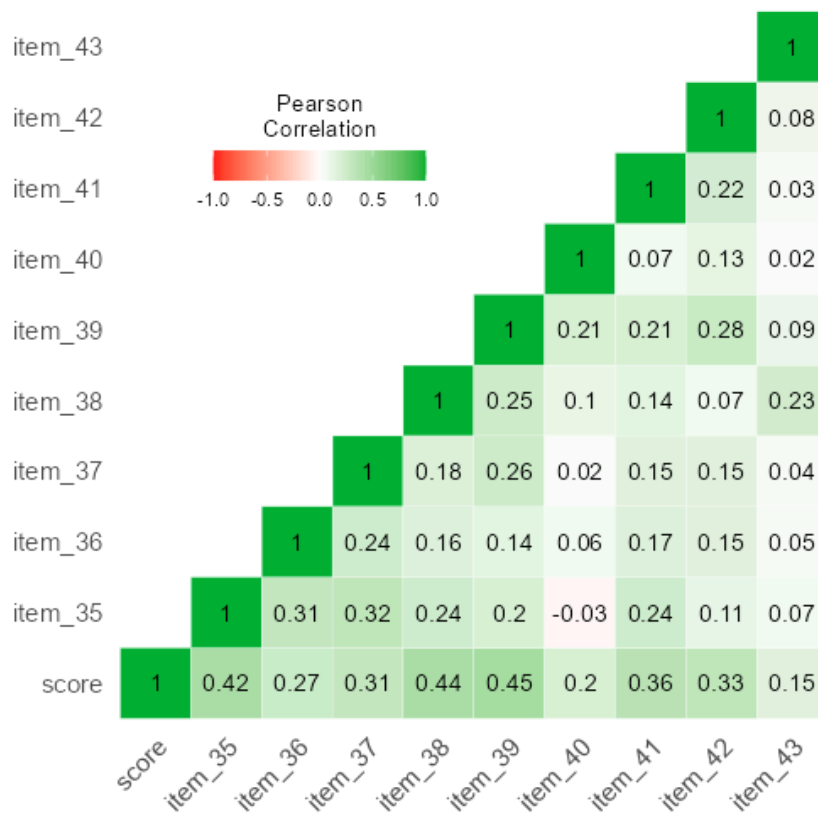
QASS Subdomain 2a



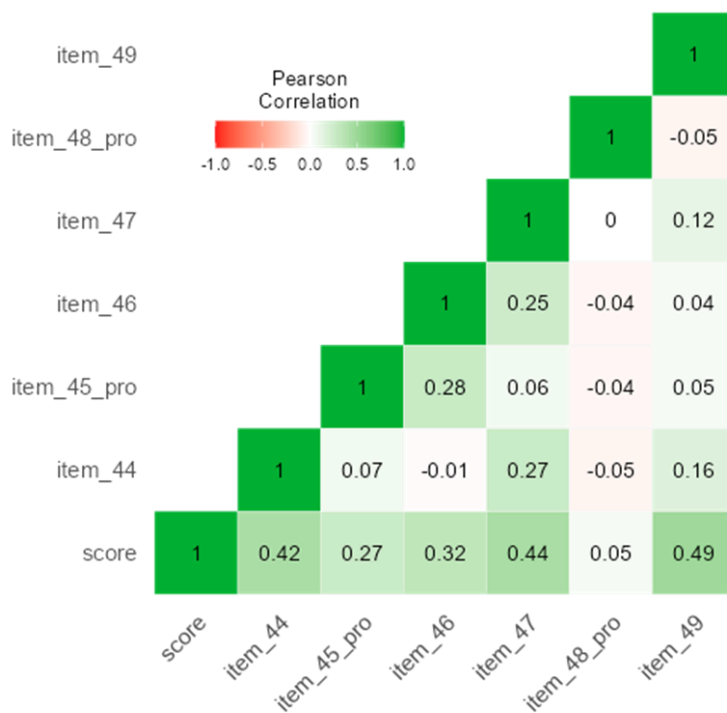
QASS Subdomain 2b



QASS Domain 3



QASS Domain 4



QASS Domain 5



REFERENCES

- Babbie, E. (2014). The practice of social research (14th ed.). CENGAGE Learning Custom Publishing.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815–824. <https://doi.org/10.1016/j.paid.2006.09.018>
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, 54, 106–148.
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kline, P. (2000). *Handbook of Psychological Testing*. London, United Kingdom: Routledge.
- Tavakol M, Wetzal A. Factor Analysis: a means for theory and instrument development in support of construct validity. *Int J Med Educ*. 2020 Nov 6;11:245-247. doi: 10.5116/ijme.5f96.0f4a. PMID: 33170146; PMCID: PMC7883798
- Giese, S., Dawes, A., Tredoux, C., Mattes, F., Bridgman, G., van der Berg, S., Schenk, J. and Kotzé J. (2022) Thrive by Five Index Report Revised August 2022, Innovation Edge, Cape Town. www.thrivebyfive.co.za.
- Henry, J. and Giese, S. (2023). Reviewing the Socio-Economic Gradient in Learning Outcomes for Children who Participated in the Thrive by Five Index, DataDrive2030, Cape Town.