




Test–retest reliability and concurrent validity of the South African Early Learning Outcomes Measure (ELOM)



Authors:

Kate J. Anderson¹ 
 Tiffany J. Henning¹ 
 Jasmin R. Moonsamy¹ 
 Megan Scott¹ 
 Christopher du Plooy² 
 Andrew R.L. Dawes^{1,3} 

Affiliations:

¹Department of Psychology, Faculty of Humanities, University of Cape Town, Cape Town, South Africa

²Division of Developmental Paediatrics, Department of Paediatrics and Child Health, Red Cross War Memorial Children's Hospital, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

³Department of International Development, University of Oxford, Oxford, United Kingdom

Corresponding author:
 Andrew Dawes,
 adkinloch1@gmail.com

Dates:

Received: 26 May 2020
 Accepted: 15 Apr. 2021
 Published: 17 June 2021

How to cite this article:

Anderson, K.J., Henning, T.J., Moonsamy, J.R., Scott, M., Du Plooy, C. & Dawes, A.R.L., 2021, 'Test–retest reliability and concurrent validity of the South African Early Learning Outcomes Measure (ELOM)', *South African Journal of Childhood Education* 11(1), a881. <https://doi.org/10.4102/sajce.v11i1.881>

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Background: The Early Learning Outcomes Measure (ELOM) assesses early learning programme outcomes in children aged 50–69 months. ELOM assesses gross motor development (GMD), fine motor coordination and visual motor integration (FMC & VMI), emergent numeracy and mathematics (ENM), cognition and executive functioning (CEF), and emergent literacy and language (ELL). Content and construct validity, reliability and cross-cultural fairness have been established.

Aim: To establish the test-retest reliability and concurrent validity of the ELOM.

Setting: Low income preschool and Grade R children.

Methods: In study one, Test-retest reliability was investigated in a convenience sample of 49 English and isiXhosa speaking preschool children (Mean age = 60.77 months, SD = 3.70) tested and retested one week apart. In study two, concurrent validity was investigated in a convenience sample of 62 children (Mean age = 75.05 months, SD = .75). ELOM performance was compared with that on the Wechsler Preschool and Primary Scale of Intelligence Fourth Edition (WPPSI-IV).

Results: Test-retest reliability was established for ELOM Total score ($r = .90, p < .001$). The concurrent validity of ELOM Total and the WPPSI-IV Full Scale Composite scores was established ($r = .64, p < .001$). FMC & VMI, CEF, and ELL domains correlated significantly with their corresponding WPPSI-IV indices: visual spatial, fluid reasoning, processing speed, working memory, and verbal comprehension.

Conclusion: The findings of both psychometric studies contribute further to the reliability and validity of the ELOM.

Keywords: concurrent validity; ELOM; WPPSI-IV; test–retest reliability; preschool.

Introduction

United Nations Sustainable Development Goal (SDG) Target 4.2. states that by 2030 countries should 'ensure that all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education' (United Nations n.d.). A key requirement of efforts to assess this outcome is the availability of reliable and valid population-level instruments suitable for children from a wide range of ethnolinguistic backgrounds, which can be used to track country attainment of SDG Goal 4.2. The Early Learning Outcomes Measure (ELOM) was developed to address the need for a locally validated, culturally fair and standardised instrument and has been used in studies of early learning programme outcomes.

As Snelling et al. (2019) note in recent years, several international efforts have been made to generate instruments to measure language, numeracy cognition and motor development in 3–5-year-old children. These include the Early Development Index (Janus 2007; <https://edi.offordcentre.com>), the International Development and Early Learning Assessment (IDELA) (Dowd et al. 2016; Pisani, Borisova & Dowd 2015; Pisani et al. 2017), the Measure of Development and Early Learning Module (MODEL) of the Measuring Early Learning Quality and Outcomes (MELQO) initiative (<http://ecdmeasure.org/about-melqo/what-is-melqo/>) and other instruments adapted to local cultural developmental settings and largely covering the same

Copyright: © 2021. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

domains as IDELA and MODEL. Examples include the East Asia-Pacific Early Child Development Scales (Rao et al. 2014) and the Tongan Early Human Capability Index (Brinkman & Vu 2016).

Limited availability of locally standardised measures adapted for multi-language and multi-cultural contexts is a challenge for research in most countries in the so-called developing world, and increasingly in the global north. With 11 official languages as well as a number of other mother tongues spoken by smaller groups, including refugees and migrants, South Africa is no exception. The ELOM direct assessment (hereafter the ELOM) was developed in response to the need for a psychometrically sound, standardised South African instrument designed to measure developmental domains associated with readiness to learn in school. Its design was informed by the South Africa's National Curriculum Framework from Birth to Four and its National Early Learning and Development Standards, which is consistent with the constructs assessed in the international early development instruments referred to above (Snelling et al. 2019). Early Learning Outcomes Measure items were drawn from reliable and valid instruments, particularly those used in Africa and other developing regions. The ELOM is a population-level instrument designed to measure the developmental status of children aged 50–69 months, which can be administered by trained non-professionals. It comprises 23 individually administered items clustered in five domains: gross motor development (GMD) measures large muscle control; fine motor coordination and visual motor integration (FMC and VMI) measure the proficiency of children's small muscle use and visual motor integration; emergent numeracy and mathematics (ENM) assesses understanding of numerical concepts, space, symbols, shapes and sizes; cognition and executive functioning (CEF) measures working memory, impulse control, problem-solving skills, critical thinking and ability to form concepts; emergent literacy and language (ELL) which assesses language use and communication skills.

Psychometric analysis has established that ELOM domains are unidimensional and internally consistent, that the instrument is reliable, and provides a fair assessment regardless of the socio-economic status (SES) or ethnolinguistic background (Snelling et al. 2019). Examination of item and domain ceiling effects on an older sample (mean age 75.82 months) compared with that used in the standardisation of the ELOM revealed that apart from three items particularly susceptible to maturation effects (one gross and two fine motor items), the remaining 20 items, four domains and ELOM total score distributions were normally distributed, or only slightly skewed (Dawes et al. 2020). Further information on the ELOM may be found at <http://elom.org.za>.

In this paper, we report on two further studies on the psychometric properties of the ELOM, which were undertaken

to complete the requirements for a psychometrically sound and reliable instrument and which have not previously been reported in the literature. These include ELOM test–retest reliability (Study 1) and ELOM concurrent validity (Study 2).

Study 1: Test–retest reliability of the Early Learning Outcomes Measure

Study 1 aimed at examining the test–retest reliability of the ELOM. The research question of interest here is whether the ELOM produces a consistent result for the same child when tested on two occasions separated by an appropriate time interval. It was hypothesised, therefore, that test scores would be significantly correlated between two administrations of the ELOM.

Research method and design

Sample

Participants were a convenience sample of English- or isiXhosa-speaking children attending two preschools that serve low-income children in Cape Town. Class lists were examined to purposively select all children in the classes who were between the ages of 55 and 69 months. G*Power 3.1.9.4 online software was used to determine sample size for correlation, which was found to be 37. After data cleaning, a sample size of $N = 49$ children ($M = 60.77$ months, standard deviation [SD] = 3.70; range 55–67 months) was realised. This sample is sufficient to detect an effect of 0.50 with power set at 0.80 ($p = 0.05$). The sample consisted of 24 male and 25 female participants, of whom 30 were English-speaking and 19 isiXhosa-speaking children.

Preparation of data for analysis

Children who were likely to show either very poor performance because of learning difficulties or invalid assessments (incomplete protocols or with scoring errors) were excluded. Once data were checked and cleaned, it was imported into Statistical Package for Social Science (SPSS) version 25.0 (IBM Corp. 2017).

Measure

Children were assessed on the ELOM as described above. The total score and scores on all five domains were used for the analysis of test–retest reliability.

Procedure

Test–retest procedure

Test–retest reliability is solely related to variability in a child's performance over time. For test–retest reliability in developmental tests such as the ELOM, having short time periods between the two assessments is recommended to ensure that the likelihood of error is because of chance and not actual changes in the child's characteristics resulting from their development (Multon 2010). Whilst a period up to 4 weeks between assessments may be acceptable for older

children and adolescents (depending on the measure), a shorter time period is recommended for preschoolers as they develop at a faster rate (Briggs-Gowan et al. 2016). WPPSI-IV test-retest intervals ranged from 7 to 48 days with an average of 23 days (Syeda & Climie 2014), whilst the testing intervals for the Early Screening Inventory were 7–10 days apart (Meisels et al. 1993). Following this pattern, the testing interval from test to re-test in the current study was 7 days.

The ELOM was administered at preschool for each child by certified ELOM assessors (<http://elom.org.za/for-assessors/>), which took approximately 45 min to an hour. To limit the likelihood of fatigue, which reduces the reliability of the assessments, all children were tested in the morning (Furr & Bacharach 2014). Assessors captured the children's information and test performance on tablets programmed to calculate the ELOM domain and total standard scores. Following each assessment, the record was uploaded to a password-protected central server and was kept confidential.

Data analysis

The Pearson product-moment correlation was used to assess the relationship between children's ELOM scores derived at the two times of measurement (Rust & Golombok 2014; Warner 2013). Study 1 drew on other studies of test-retest reliability with similar instruments to set a criterion for an acceptable correlation between the scores derived at the two points of measurement. Bryant and Roffe (1978) reported the test-retest reliability (Pearson's r) of the McCarthy Scales to range from 0.71 to 0.85. As the WPPSI-IV and the ELOM are more comparable instruments (see Study 2), we followed Syeda and Climie (2014) in setting the criterion for an acceptable ELOM test-retest reliability coefficient at 0.75.

Ethical considerations

The study was approved by the University of Cape Town's Humanities Faculty Ethics Committee (PSY2019-024). Participating preschool staff were briefed on the study. Prior to testing, parents or guardians of the participating children were requested to give written informed consent for their child's participation by signing an informed consent form. As there was a high likelihood of parents forgetting to return the forms to school with the child, passive consent was used where necessary (as approved by the Ethics Committee). Children were informed that they could stop the assessment at any time without consequences and could also request for a break during the assessment.

Results

As normality was violated and linearity was fairly but weakly upheld, the data were bootstrapped and confidence intervals were established (Field 2013; how2stats 2019; Swank & Mullen 2017). As is evident (Table 1), the ELOM total score (0.90), FMC and VMI (0.79), ENM (0.76) and arguably ELL (0.74) either exceeded or met the criterion

chosen. Cognition and executive functioning (0.64) and GMD (0.50) were below the criterion. The ELOM total score test-retest reliability exceeded the level (0.80) put forward for group-level analysis by Cronbach (Polit 2014). None of the confidence intervals crossed zero and were narrow with a difference of less than 0.4 (Cumming 2012). All p values were considered to be statistically significant at $p < 0.001$.

Discussion

In this study population, using ELOM total scores, the instrument has an excellent test-retest reliability (0.90) over a 7-day period. This finding is in line with the test-retest reliability of the WPPSI-IV Full Scale IQ (0.93) and its composite scores (0.84–0.89) reported by Syeda and Climie (2014), and is in line with coefficients of 0.82–0.92 for the same WPPSI-IV composites reported by Soares and McCrimmon (2013). The FMC and VMI, ENM domains and arguably ELL met the criterion for acceptable test-retest reliability chosen for this study.

Study 1 has limitations. A convenience sample was used as random sampling was not practical in the two schools from which the children were drawn (all children of the appropriate age had to be included to make up the sample). In addition, the study sample was drawn from children in lower socio-economic groups. It is possible, although very unlikely, that the test-retest reliability of the ELOM could differ in a study of children from higher SES backgrounds. This is because this form of reliability is a property of the test and not the population. As Aldridge, Dovey and Wade (2017) stated, test-retest reliability:

[R]efers to the systematic examination of consistency, reproducibility, and agreement among two or more measurements of the same individual, using the same tool, under the same conditions (i.e. when we don't expect the individual being measured to have changed on the given outcome). Test-retest studies help us to understand how dependable our measurement tools are likely to be if they are put into wider use in research. (p. 208)

As noted, further studies of the test-retest reliability of the ELOM should be conducted with random samples and in children from higher socio-economic backgrounds and other language groups so as to ensure that the results reported here are confirmed.

TABLE 1: Test-retest reliability coefficients for the Early Learning Outcomes Measure.

ELOM domains and total	Test-retest reliability		
	r	p	95% CI
Gross motor development	0.50	< 0.001	0.29, 0.68
Fine motor coordination and visual motor integration	0.79	< 0.001	0.63, 0.89
Emergent numeracy and mathematics	0.76	< 0.001	0.57, 0.89
Cognition and executive functioning	0.64	< 0.001	0.44, 0.83
Emergent language and literacy	0.74	< 0.001	0.58, 0.86
ELOM total	0.90	< 0.001	0.83, 0.95

ELOM, early learning outcomes measure.

Study 2: Concurrent validity of the Early Learning Outcomes Measure

The primary aim of Study 2 was to establish the concurrent validity of the ELOM by comparing children's performance on the instrument with core subtests of the WPPSI-IV that measure the same constructs. We investigated whether concurrent validity was demonstrated between ELOM total and WPPSI-IV Full Scale composite scores, and between three selected ELOM domains (FMC and VMI, CEF and ELL) and WPPSI-IV indices (visual spatial, fluid reasoning, processing speed, working memory and verbal comprehension). Study 2 aimed at making a contribution to the psychometric qualities of the ELOM by strengthening its validity. The establishment of concurrent validity would mean that ELOM results can be interpreted with greater confidence, and thus, with wider application and relevance.

Research method and design

Sample

Participants were already enrolled in the Drakenstein Child Health Study (DCHS), birth cohort study being conducted in Paarl in the Western Cape of South Africa that follows 1000 mother-child dyads from 20–28 weeks' gestation. The DCHS participants are all of low SES and are vulnerable to substance abuse and human immunodeficiency virus (HIV) (Stein et al. 2015). G*Power 3.1.9.4 online software was used to determine sample size for correlation, with power set to 0.80, an effect size of 0.40 and significance set to 0.05, for a one-tailed test. These requirements yielded a minimum required sample size of 37 (Faul et al. 2009). After cleaning, the sample size of $N = 62$ (24 male and 38 female participants) provided sufficient statistical power (>0.80) to accurately assess concurrent validity. The age range of sample was from 72.98 to 75.97 months ($M = 75.05$, $SD = 0.75$). This included 45 isiXhosa, 16 Afrikaans and one English speaker. The demographic characteristics of the whole DCHS sample are provided in Stein et al. (2015). These children were older than the ELOM standardisation range (50–69 months). However, as noted above, ceiling effects in this age group are only evident for three of the 23 ELOM items and are not evident for ELOM

total and domain scores. It was, therefore, decided that the ELOM could be used in this age group to investigate concurrent validity.

Preparation of data for analysis

As for Study 1, ELOM Direct Assessment guidelines were used to exclude records of children likely to show either very poor performance because of learning difficulties or invalid assessments. Once data were cleaned, it was imported into SPSS version 25.0 (IBM Corp. 2017).

Measures

Participants for Study 2 were tested on the ELOM (described above) and the WPPSI-IV core subtests during the 72-month neurocognitive DCHS testing wave in 2019. The ELOM total scores, and scores on three selected domains (FMC & VMI, CEF and ELL), which are related to areas of the WPPSI-IV core subtests, were used to assess the concurrent validity. The WPPSI-IV is a standardised intelligence test used for children between 30 and 91 months (Wechsler 2012a), which has not been standardised for use in South Africa. Strong test-retest reliability and concurrent validity have been established (Thorndike 2014). The WPPSI-IV Full Scale composite score is comprised of five Primary Index Scales: verbal comprehension, visual spatial, fluid reasoning, working memory and processing speed indices. The children in this DCHS are tested on WPPSI-IV core subtests: Information, Similarities, Block Design, Matrix Reasoning, Picture Memory and Bug Search (see Table 2). These contribute to the indices, which combine to derive Full-Scale IQ (the WPPSI-IV Full Scale composite score). The core WPPSI-IV subtests were compared, via WPPSI-IV indices, with ELOM domains (Table 3).

Procedure

Ethical considerations

The Faculty of Health Sciences, Human Research Ethics Committee at the University of Cape Town (401/2009) and the Western Cape Provincial Health Research Committee (2011RP45) approved the DCHS (including ELOM and WPPSI-IV administration).

TABLE 2: Details of WPPSI-IV core subtests.

WPPSI-IV subtest	WPPSI-IV index	Explanation of subtest	Subtest constructs
Information	Verbal comprehension index	Through verbal and picture items, the child responds to general knowledge questions.	Acquisition, retention and retrieval of factual knowledge from long-term memory (Groth-Marnat 2003; Wechsler 2012b).
Similarities	Verbal comprehension index	Children identify and explain why groups or pairs of objects or words are similar.	Abstract reasoning and verbal fluency (Groth-Marnat 2003; Wechsler 2012b).
Block design	Visual spatial index	Children use blocks to redesign a stimulus picture.	Visual motor coordination and non-verbal problem-solving abilities (Groth-Marnat 2003; Wechsler 2012b).
Matrix reasoning	Fluid reasoning index	The child chooses the response option that best completes an incomplete matrix.	Fluid intelligence and the ability to analyse the relationship between a whole and its parts (Groth-Marnat 2003; Wechsler 2012b).
Picture memory	Working memory index	The child is shown a picture and must remember the stimulus by choosing it out of a set of response options.	Working memory using proactive interference (Canivez 2014; Wechsler 2012b).
Bug search	Processing speed index	Children choose out of an array of insects, the one that matches the target insect.	Processing speed of visual information, attention and concentration (Groth-Marnat 2003; Wechsler 2012b).

WPPSI-IV, wechsler preschool and primary scale of intelligence fourth edition.

TABLE 3: Early Learning Outcomes Measure and WPPSI-IV comparison.

WPPSI core subtest	WPPSI index	ELOM domain	Justification for link between the WPPSI index and ELOM domain
Block design	Visual spatial	Fine motor coordination and visual motor integration	Visual motor coordination and visual spatial integration are important aspects of fine motor skills (Carlson, Rowe & Curby 2013; Decker et al. 2011).
Matrix reasoning	Fluid reasoning	Cognition and executive functioning	Fluid reasoning, processing speed and working memory are well-established dimensions of executive functioning (Anderson 2002; Brocki & Bohlin 2004; Decker, Hill & Dean 2007; Karasinski 2015; Salthouse 2005).
Bug search	Processing speed	Cognition and executive functioning	
Picture memory	Working memory	Cognition and executive functioning	
Similarities	Verbal comprehension	Emergent literacy and language	Verbal comprehension and verbal fluency (both aspects of the VCI) are components of language (Maseda et al. 2014).
Information	Verbal comprehension	Emergent literacy and language	

ELOM, early learning outcomes measure; WPPSI, wechsler preschool and primary scale of intelligence fourth edition; VCI, verbal comprehension index.

Concurrent validity procedure

The assessors for the DCHS with postgraduate psychology qualifications administered both tests to children in private rooms at study sites. The ELOM was administered in the child's home language (as the instrument is available in Afrikaans and isiXhosa). As the test has not been translated into South African languages, the WPPSI-IV was administered in English with translation into isiXhosa or Afrikaans by an assistant during the testing sessions. In order to reduce the likelihood of variation in translations, the DCHS devised standard translations for use by all assistants. All WPPSI-IV translations were forward and back translated; thereafter, translation consensus meetings were carried out with community nursing staff and the translators to ensure that the translations were age and context appropriate.

Both instruments were administered on the same day, with the ELOM first and then the WPPSI-IV. Children were given a break between the two testing sessions.

Data analysis

Pearson's correlation coefficient (r) was used to measure the strength of relationships between WPPSI-IV core subtests, WPPSI-IV indices, ELOM items and ELOM domains. The criteria for acceptable r (see Table 4) were followed according to Swank and Mullen (2017) who noted that correlation coefficients used in testing validity are lower than other applications of correlation, as abstract or latent constructs result in measurement complexities.

Results

Descriptive statistics for ELOM and WPPSI-IV scores respectively were displayed (Table 5 and Table 6). The correlations between ELOM and WPPSI-IV scores are provided (Table 7). The very high correlation ($r = 0.64$; $p < 0.001$) between the ELOM total Score and the WPPSI-IV Full Scale composite score demonstrates a strong concurrent validity. All the three ELOM domains yielded a high or very high correlation with the WPPSI-IV Full Scale composite score ($p < 0.001$). The expected correlations from Table 3 are highlighted (Table 8) and it shows the strongest relationships existing between ELOM domains and WPPSI-IV subtests. A significant correlation was observed when the ELOM items were individually correlated with the WPPSI-IV core subtests, with results shown (Table 9).

TABLE 4: Categorisation criteria for Pearson coefficients.

r^\dagger	Correlation strength
< 0.20	Low
$\leq 0.20 \times < 0.40$	Moderate
$\leq 0.40 \times < 0.50$	High
≥ 0.50	Very high

† , r represents the Pearson correlation statistic, in *Rho*.

Discussion

Strong concurrent validity of the ELOM with the WPPSI-IV has been established in this sample. Both tests measure similar constructs. The very high and significant correlation between the ELOM total score and the WPPSI-IV Full Scale composite score suggests that the ELOM total score could be used as a proxy indicator of IQ, particularly as the ELOM is standardised for South Africa, whereas the WPPSI-IV is not. However, investigation of the relationship between the two tests in children from across a wide range of socio-economic backgrounds is necessary before this can be confirmed. The FMC & VMI domain showed the strongest correlation with WPPSI-IV Bug Search, suggesting that they are measuring similar constructs – perhaps a visual aspect. The CEF domain showed the strongest correlation with WPPSI-IV Block Design, suggesting that they are measuring similar constructs – potentially non-verbal problem solving and spatial perception (Groth-Marnat 2003; Wechsler 2012b). As expected, the ELL domain showed the strongest correlation with the WPPSI-IV VCI composite score (see Table 3).

A limitation of Study 2 is that the sample was 6 months older than the ELOM standardisation sample and that all children were from low socio-economic backgrounds, as the DCHS tracks the development of children growing up in high-risk circumstances (Stein et al. 2015). Replication with children from the full range of socio-economic backgrounds is recommended.

Conclusion

The ELOM was developed because of the lack of standardised instruments in South Africa suitable for measuring early learning programme effects and children's readiness to learn in the Grade R year (Snelling et al. 2019). It is the first psychometrically robust population-level South African instrument that can be administered by trained non-professionals at low cost, which is used to assess preschool children from across a wide range of socio-economic and ethnolinguistic backgrounds. Prior to the current studies,

TABLE 5: Early Learning Outcomes Measure descriptive statistics.

Variable	N	Min	Max	Mean	SD	Skewness	SE	Kurtosis	SE**
ELOM total score	62	31.86	85.45	60.51	11.55	-0.04	0.30	-0.57	0.60
FMC and VMI domain score	62	10.73	20.00	16.92	2.47	-0.68	0.30*	-0.54	0.60
CEF domain score	62	2.34	18.08	10.50	3.77	0.17	0.30	-0.65	0.60
ELL domain score	62	2.19	19.27	9.93	4.40	0.26	0.30	-0.55	0.60
Valid N	62	-	-	-	-	-	-	-	-

ELOM, early learning outcomes measure; FMC, fine motor coordination; VMI, visual motor integration; CEF, cognition and executive functioning; ELL, emergent literacy and language; SD, standard deviation; SE, standard error.

*, FMC and VMI domain is moderately skewed, but all other skewness statistics are acceptable.

** , All kurtosis statistics are acceptable.

TABLE 6: WPPSI-IV descriptive statistics.

Variable	N	Min	Max	Mean	SD	Skewness	SE	Kurtosis	SE
Full scale composite score	62	56.00	88.00	71.53	8.36	0.16	0.30	-0.86	0.60
VCI composite score	62	54.00	95.00	73.34	10.21	0.20	0.30	-0.59	0.60
Block design scaled score	62	3.00	13.00	7.18	1.51	0.93	0.30*	4.30	0.60*
Matrix reasoning scaled score	62	1.00	8.00	4.68	2.19	-0.07	0.30	-0.99	0.60
Bug search scaled score	62	2.00	12.00	7.15	2.31	0.18	0.30	-0.11	0.60
Picture memory scaled score	62	3.00	11.00	7.27	1.82	-0.30	0.30	-0.50	0.60
Similarities scaled score	62	1.00	10.00	5.11	2.32	0.31	0.30	-0.82	0.60
Information scaled score	62	2.00	9.00	5.37	1.98	-0.17	0.30	-1.25	0.60
Valid N	62	-	-	-	-	-	-	-	-

VCI, xxx; SD, standard deviation; SE, standard error.

* , Block design scaled score is moderately skewed, but all other skewness and kurtosis statistics are acceptable.

TABLE 7: Correlations of Early Learning Outcomes Measure domains and WPPSI-IV subtests.

Variable	WPPSI-IV subtests							VCI composite score	Full Scale composite score	
	Block Design	Matrix Reasoning (MR)	Bug Search (BS)	Picture Memory (PM)	Similarities	Information				
ELOM Domain	FMC & VMI	r^{\dagger}	0.34**	0.39**	0.51***	0.38**	0.22*	0.18	0.25*	0.53***
		p	0.003	0.001	< 0.001	0.001	0.046	0.078	0.025	< 0.001
		CI [LL, UL]‡	0.10, 0.54	0.15, 0.58	0.30, 0.68	0.15, 0.58	-0.04, 0.44§	-0.07, 0.41§	0.001, 0.47	0.32, 0.69
CEF		r^{\dagger}	0.37**	0.35**	0.35**	0.32**	0.22*	0.29*	0.32**	0.49***
		p	0.002	0.003	0.003	0.005	0.041	0.012	0.006	< 0.001
		CI [LL, UL]‡	0.13, 0.57	0.11, 0.55	0.11, 0.55	0.08, 0.53	-0.03, 0.45§	0.04, 0.50	0.07, 0.52	0.28, 0.66
ELL		r^{\dagger}	0.25*	0.17	0.44***	0.37**	0.46***	0.32**	0.50***	0.54***
		p	0.027	0.093	< 0.001	0.001	< 0.001	0.006	< 0.001	< 0.001
		CI [LL, UL]‡	-0.004, 0.47§	-0.08, 0.40§	0.22, 0.62	0.14, 0.57	0.24, 0.64	0.08, 0.53	0.28, 0.66	0.33, 0.70
ELOM total score		r^{\dagger}	0.43***	0.32**	0.48***	0.53***	0.39**	0.29*	0.43***	0.64***
		p	< 0.001	0.006	< 0.001	< 0.001	0.001	0.011	< 0.001	< 0.001
		CI [LL, UL]‡	0.20, 0.61	0.07, 0.52	0.26, 0.65	0.33, 0.69	0.16, 0.59	0.05, 0.51	0.20, 0.62	0.47, 0.77

ELOM, early learning outcomes measure; FMC, fine motor coordination; VMI, visual motor integration; CEF, cognition and executive functioning; ELL, emergent literacy and language; CI, confidence interval.

†, r represents the Pearson correlation statistic, in Rho .

‡, CI stands for confidence interval, calculated at the 95% interval. LL stands for lower limit, and UL stands for upper limit.

§, It is noted that these confidence intervals span across zero, so these correlations should be interpreted with caution.

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

TABLE 8: Expected versus strongest relationships of Early Learning Outcomes Measure domains with WPPSI-IV subtests.

ELOM	Expected relationship as per Table 3			Strongest relationship		
	WPPSI-IV	r^{\dagger}	Strength of relationship	WPPSI-IV	r	Strength of relationship
ELOM total score	WPPSI-IV full scale composite score	0.64***	Very high	‡	-	-
FMC and VMI	Block design	0.34**	Moderate	Bug Search	0.51***	Very high
CEF	Matrix reasoning	0.35**	Moderate	Block design	0.37**	-
	Bug search	0.35**	Moderate	-	-	Moderate
	Picture memory	0.32**	Moderate	-	-	-
ELL	Verbal comprehension index	0.50***	Very high	‡	-	-

ELOM, early learning outcomes measure; FMC, fine motor coordination; VMI, visual motor integration; CEF, cognition and executive functioning; ELL, emergent literacy and language.

†, r represents the Pearson correlation statistic, in Rho .

‡, In the case of the ELOM total score and the ELL domain, the expected relationship was also the strongest.

, $p < 0.01$; *, $p < 0.001$.

TABLE 9: Correlations of ELOM items and WPPSI-IV subtests.

ELOM domain§	ELOM Item	WPPSI-IV subtests						
		Block design	Matrix reasoning	Bug search	Picture memory	Similarities	Information	
Gross Motor Development	1	r_{\dagger}^{\dagger}	-0.12	-0.12	0.02	0.19	-0.14	0.14
		p	0.177	0.175	0.445	0.065	0.140	0.135
		CI [LL, UL]	-0.40, 0.13†	-0.36, 0.13†	-0.23, 0.27†	-0.06, 0.42†	-0.38, 0.11†	-0.11, 0.38†
	2	r_{\dagger}^{\dagger}	0.10	-0.31**	-0.10	0.04	-0.14	-0.27*
		p	0.215	0.008	0.233	0.393	0.136	0.018
		CI [LL, UL]	-0.15, 0.34†	-0.52, -0.06	-0.34, 0.16†	-0.22, 0.28†	-0.38, 0.11†	-0.48, -0.02
	3	r_{\dagger}^{\dagger}	0.08	-0.01	-0.16	-0.02	-0.18	-0.05
		p	0.277	0.456	0.113	0.427	0.083	0.358
		CI [LL, UL]	-0.18, 0.32†	-0.26, 0.24†	-0.39, 0.10†	-0.27, 0.23†	-0.41, 0.08†	-0.29, 21†
	4	r_{\dagger}^{\dagger}	0.15	0.04	-0.08	0.21*	-0.008	-0.10
		p	0.127	0.381	0.263	0.049	0.275	0.229
		CI [LL, UL]	-0.11, 0.38†	-0.21, 0.29†	-0.33, 0.17†	-0.04, 0.44†	-0.32, 0.18†	-0.34, 0.16†
FMC and VMI	5	r_{\dagger}^{\dagger}	0.30**	0.25*	0.21	0.23*	0.03	0.18
		p	0.008	0.026	0.053	0.036	0.425	0.086
		CI [LL, UL]	0.06, 0.51	-0.002, 0.47†	-0.04, 0.44†	-0.02, 0.45†	-0.23, 0.27†	-0.08, 0.41†
	6	r_{\dagger}^{\dagger}	0.17	0.35**	0.32**	0.34**	0.11	0.18
		p	0.097	0.003	0.006	0.004	0.198	0.077
		CI [LL, UL]	-0.09, 0.40†	0.11, 0.55	0.07, 0.52	0.10, 0.54	-0.14, 0.35†	-0.07, 0.41†
	7	r_{\dagger}^{\dagger}	0.01	-0.12	0.01	0.11	-0.01	-0.32**
		p	0.464	0.178	0.460	0.203	0.481	0.006
		CI [LL, UL]	-0.24, 0.26†	-0.40, 0.13†	-0.24, 0.26†	-0.15, 0.35†	-0.26, 0.24†	-0.53, -0.08
	8	r_{\dagger}^{\dagger}	0.17	0.12	0.36**	0.10	0.18	0.06
		p	0.099	0.186	0.002	0.231	0.079	0.315
		CI [LL, UL]	-0.09, 0.40†	-0.14, 0.36†	0.12, 0.56†	-0.16, 0.34†	-0.07, 0.41†	-0.20, 0.31†
Emergent Numeracy and Mathematics	9	r_{\dagger}^{\dagger}	0.24*	0.12	0.30*	0.20	0.38**	0.25*
		p	0.031	0.172	0.010	0.060	0.001	0.026
		CI [LL, UL]	-0.01, 0.46†	-0.13, 0.36†	0.05, 0.51	-0.05, 0.43†	0.15, 0.58	-0.002, 0.47†
	10	r_{\dagger}^{\dagger}	0.17	0.06	0.15	0.25*	0.19	0.15
		p	0.117	0.336	0.149	0.039	0.093	0.150
		CI [LL, UL]	-0.08, 0.40†	-0.19, 0.31†	-0.10, 0.39†	0.002, 0.47	-0.06, 0.42†	-0.10, 0.38†
	11	r_{\dagger}^{\dagger}	0.19	0.25*	0.18	0.32**	0.30**	0.30**
		p	0.075	0.028	0.085	0.006	0.008	0.009
		CI [LL, UL]	-0.07, 0.42†	-0.01, 0.47†	-0.08, 0.41†	0.08, 0.53	0.06, 0.51	0.05, 0.51
	12	r_{\dagger}^{\dagger}	0.003	0.01	0.13	0.09	0.09	-0.09
		p	0.492	0.479	0.152	0.238	0.247	0.246
		CI [LL, UL]	-0.25, 0.25†	-0.24, 0.26†	-0.12, 0.37†	-0.16, 0.33†	-0.17, 0.33†	-0.33, 0.16†
13	r_{\dagger}^{\dagger}	-0.01	0.13	0.24*	0.27*	0.24*	0.02	
	p	0.459	0.151	0.032	0.016	0.028	0.428	
	CI [LL, UL]	-0.26, 0.24†	-0.12, 0.37†	-0.02, 0.46†	0.03, 0.49	-0.01, 0.47†	-0.23, 0.27†	
CEF	14	r_{\dagger}^{\dagger}	0.14	-0.02	0.18	0.05	0.10	0.11
		p	0.137	0.450	0.079	0.344	0.225	0.208
		CI [LL, UL]	-0.11, 0.38†	-0.27, 0.24†	-0.07, 0.41†	-0.20, 0.30†	-0.16, 0.34†	-0.15, 0.35†
	15	r_{\dagger}^{\dagger}	0.31**	0.38**	0.13	0.20	0.20	0.30*
		p	0.008	0.002	0.166	0.069	0.063	0.010
		CI [LL, UL]	0.07, 0.52	0.14, 0.57	-0.13, 0.37†	-0.06, 0.42†	-0.05, 0.43†	0.06, 0.51
	16	r_{\dagger}^{\dagger}	-0.12	0.12	0.18	0.16	0.01	0.04
		p	0.182	0.192	0.087	0.118	0.463	0.374
		CI [LL, UL]	-0.36, 0.13†	-0.14, 0.36†	-0.07, 0.41†	-0.10, 0.39†	-0.24, 0.26†	-0.21, 0.29†
	17	r_{\dagger}^{\dagger}	0.41**	0.29*	0.31**	0.21*	0.18	0.22*
		p	0.001	0.011	0.007	0.049	0.083	0.042
		CI [LL, UL]	0.17, 0.60	0.04, 0.50	0.07, 0.52	-0.04, 0.44†	-0.08, 0.41†	-0.03, 0.45†
ELL	18	r_{\dagger}^{\dagger}	0.15	0.01	0.07	0.20	0.25*	0.14
		p	0.118	0.486	0.300	0.059	0.027	0.138
		CI [LL, UL]	-0.10, 0.40†	-0.25, 0.25†	-0.19, 0.31†	-0.05, 0.43†	-0.01, 0.47†	-0.11, 0.38†
	19	r_{\dagger}^{\dagger}	0.17	0.16	0.48***	0.16	0.20	0.22*
		p	0.094	0.108	<0.001	0.115	0.058	0.046
		CI [LL, UL]	-0.08, 0.40†	-0.09, 0.39†	0.27, 0.65	-0.10, 0.40†	-0.05, 0.43†	-0.04, 0.44†

Table 9 continues on next page→

TABLE 9 (continues...): Correlations of ELOM items and WPPSI-IV subtests.

ELOM domain§	ELOM Item	WPPSI-IV subtests					
		Block design	Matrix reasoning	Bug search	Picture memory	Similarities	Information
20	r_{\dagger}^{\ddagger}	0.12	0.06	0.27*	0.16	0.25*	0.20
	ρ	0.186	0.328	0.017	0.105	0.025	0.059
	CI [LL, UL]	-0.14, 0.36†	-0.20, 0.30†	0.02, 0.49	-0.09, 0.40†	0.002, 0.47	-0.05, 0.43†
21	r_{\dagger}^{\ddagger}	0.08	-0.03	0.24*	0.15	0.30**	0.34**
	ρ	0.261	0.417	0.030	0.121	0.008	0.003
	CI [LL, UL]	-0.17, 0.33†	-0.28, 0.22†	-0.01, 0.46†	-0.10, 0.39†	0.06, 0.51	0.10, 0.54
22	r_{\dagger}^{\ddagger}	0.05	0.32**	0.16	0.23*	0.48***	0.17
	ρ	0.353	0.006	0.102	0.035	< 0.001	0.088
	CI [LL, UL]	-0.20, 0.30†	0.08, 0.53	-0.09, 0.40†	-0.02, 0.46†	0.26, 0.65	-0.08, 0.41†
23	r_{\dagger}^{\ddagger}	0.36**	0.27*	0.40**	0.45**	0.61***	0.24
	ρ	0.008	0.041	0.004	0.001	< 0.001	0.060
	CI [LL, UL]	0.13, 0.56	0.02, 0.49	0.17, 0.59	0.23, 0.63	0.42, 0.75	-0.01, 0.46†

ELOM, early learning outcomes measure; FMC, fine motor coordination; VMI, visual motor integration; CEF, cognition and executive functioning; ELL, emergent literacy and language; CI, confidence interval; LL, lower limit; UL, upper limit.

†, It is noted that these confidence intervals span across zero, so these correlations should be interpreted with caution.

‡, r represents the Pearson correlation statistics, in *Rho*.

§, Item-by-item analyses of the gross motor development domain and the emergent numeracy and mathematics domain were excluded from this study as these domains do not fall into the scope of this research.

*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$.

test-retest reliability and concurrent validity had not been established. Whilst the concurrent validity of the GMD and the ENM domains of the ELOM remain to be established, these studies have enhanced the psychometric properties of the measure.

Acknowledgements

For Study 1, the authors are grateful to Innovation Edge (innovationedge.org) for funding. They are also grateful to the ELOM project manager for sourcing study participants, to the ELOM data manager for providing the data, to the ELOM assessors for their careful test administration and to the preschools for permitting them to use their facilities. With regard to Study 2, they are most grateful to the principal investigators of the Drakenstein Child Health Study (DCHS), and in particular, to Prof. Kirsty Donald, for granting them permission to use their data for psychometric analysis.

Competing interests

The authors have declared that no competing interests exist.

Authors' contributions

T.J.H. and J.R.M. collected the data, contributed equally to the analysis of data, and wrote the findings for Study 1 – Test-retest reliability. K.J.A. and M.S. contributed equally to the analysis of data and write up of findings for Study 2 – concurrent validity. C.d.P. managed the collection of data for Study 2. A.R.L.D. designed both studies and contributed to the write up of the article as a whole. All authors approved for the submission of the manuscript.

Funding information

The research study was entirely funded by Innovation Edge: innovationedge.org.za

Data availability

Data from Study 1 can be shared – Application must be made to the corresponding author. Data for Study 2 cannot be shared as it forms part of a longitudinal study that is still in progress, and the data are not yet provided in the public domain.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References

- Aldridge, V.K., Dovey, T.M. & Wade, A., 2017, 'Assessing test-retest reliability of psychological measures', *European Psychologist* 22(4), 207–218. <https://doi.org/10.1027/1016-9040/a000298>
- Anderson, P., 2002, 'Assessment and development of executive function (EF) during childhood', *Child Neuropsychology* 8(2), 71–82. <https://doi.org/10.1076/chin.8.2.71.8724>
- Briggs-Gowan, M.J., Godoy, L., Heberle, A. & Carter, A.S., 2016, 'Assessment of psychopathology in young children', in D. Cicchetti (ed.), *Developmental psychopathology, theory and method*, pp. 1–45, viewed 19 July 2019, from <https://books.google.co.za/books?id=ENE9CgAAQBAJ&pg=PA13&dq=test-retest+developmental&hl=en&sa=X&ved=0ahUKewip4eqN34bhAhXFRBUIHWgID3EQ6AEIKDAA#v=onepage&q=test-retest%20developmental&f=false>.
- Brinkman, S. & Vu, B.T., 2016, *Early childhood development in Tonga: Baseline results from the Tongan early human capability index*. World Bank Publications, Washington, DC, viewed 20 May 2019, from <https://openknowledge.worldbank.org/handle/10986/25674>.
- Brocki, K.C. & Bohlin, G., 2004, 'Executive functions in children aged 6 to 13: A dimensional and developmental study', *Developmental Neuropsychology* 26(2), 571–593. https://doi.org/10.1207/s15326942dn2602_3
- Bryant, C.K. & Roffe, M.W., 1978, 'A reliability study of the McCarthy scales of children's abilities', *Journal of Clinical Psychology* 34(2), 401–406. [https://doi.org/10.1002/1097-4679\(197804\)34:2%3C401::AID-JCLP2270340230%3E3.0.CO;2-V](https://doi.org/10.1002/1097-4679(197804)34:2%3C401::AID-JCLP2270340230%3E3.0.CO;2-V)
- Canivez, G.L., 2014, 'Test review of Wechsler preschool and primary scale of intelligence – Fourth edition', in J.F. Carlson, K.F. Geisinger & J.L. Jonson (eds.), *The nineteenth mental measurements yearbook*, pp. 4–15, Buros Center for Testing, Lincoln Nebraska, NE.
- Carlson, A.G., Rowe, E. & Curby, T.W., 2013, 'Disentangling fine motor skills' relations to academic achievement: The relative contributions of visual spatial integration and visual motor coordination', *The Journal of Genetic Psychology* 174(5), 514–533. <https://doi.org/10.1080/00221325.2012.717122>
- Cumming, G., 2012, 'Chapter 3: Confidence intervals', in *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*, pp. 53–86, viewed 05 August 2019, from https://books.google.co.za/books?id=fgqR3aBtuVkc&printsec=frontcover&dq=confidence+interval+expand+zero+andy+field&hl=en&sa=X&ved=0ahUKewjG-5eMu4DIhWzmFwKHXg_CTKQ6AEITDAF#v=snippet&q=confidence%20&f=false.

- Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L. & Tredoux, C.G., 2020, *Early learning outcomes measure*, Technical manual, 3rd edn., The Innovation Edge, Claremont, Cape Town, viewed 10 September 2020, from http://elom.org.za/wp-content/uploads/2020/06/ELOM-Technical-Manual_2020.pdf.
- Decker, S.L., Englund, J.A., Carboni, J.A. & Brooks, J.H., 2011, 'Cognitive and developmental influences in visual motor integration skills in young children', *Psychological Assessment* 23(4), 1010–1016. <https://doi.org/10.1037/a0024079>
- Decker, S.L., Hill, S.K. & Dean, R.S., 2007, 'Evidence of construct similarity in executive functions and fluid reasoning abilities', *International Journal of Neuroscience* 117, 735–748. <https://doi.org/10.1080/00207450600910085>
- Dowd, A.J., Borisova, I., Amente, A. & Yenew, A., 2016, 'Realizing capabilities in Ethiopia: Maximizing early childhood investment for impact and equity', *Journal of Human Development and Capabilities* 17(4), 477–493. <https://doi.org/10.1080/19452829.2016.1225702>
- Faul, F., Erdfelder, E., Buchner, A. & Lang, A.G., 2009, 'Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses', *Behavior Research Methods* 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Field, A., 2013, 'Correlation', in *Discovering statistics using IBM SPSS statistics*, 4th edn., pp. 262–292, viewed 08 August 2019, from https://books.google.co.za/books?hl=en&lr=&id=c0Wk9luBmAoC&oi=fnd&pg=PP2&dq=andy+field+2013+ibm&ots=LbFqKN0x5D&sig=OhxCLh_9cbpbHzdd0HmJTe2AEBg&redir_esc=y#v=onepage&q=With%20bivariate%20analysis&f=false.
- Furr, R.M. & Bacharach, V.R., 2014, *Psychometrics: An introduction*, Sage, Thousand Oaks, CA.
- Groth-Marnat, G., 2003, *The handbook of psychological assessment*, John Wiley & Sons, Inc., New York, NJ.
- how2stats, 2019, *Bootstrapping bug in SPSS – Fix*, viewed 12 August 2019, from https://www.youtube.com/watch?v=ut_C8OhHeec.
- IBM Corp, 2017, IBM SPSS Statistics for Windows [Computer software], IBM Corp, Armonk, NY.
- Janus, M.E., 2007, 'The early development instrument: A tool for monitoring children's development and readiness for school', in M.E. Young & L.M. Richardson (eds.), *Early child development from measurement to action: A priority for growth and equity*, pp. 183–202, World Bank Publications, Washington, DC.
- Karasinski, C., 2015, 'Language ability, executive functioning and behaviour in school-age children', *International Journal of Language and Communication Disorders* 50(2), 144–150. <https://doi.org/10.1111/1460-6984.12104>
- Maseda, A., Lodeiro-Fernandez, L., Lorenzo-Lopez, L., Nunez-Naveira, L., Balo, A. & Millan-Calenti, J.C., 2014, 'Verbal fluency, naming and verbal comprehension: Three aspects of language as predictors of cognitive impairment', *Aging and Mental Health* 18(8), 1037–1045. <https://doi.org/10.1080/13607863.2014.908457>
- Meisels, S.J., Henderson, L.W., Liaw, F.R., Browning, K. & Ten Have, T., 1993, 'New evidence for the effectiveness of the early screening inventory', *Early Childhood Research Quarterly* 8(3), 327–346. [https://doi.org/10.1016/S0885-2006\(05\)80071-7](https://doi.org/10.1016/S0885-2006(05)80071-7)
- Multon, K.D., 2010, 'Test-retest reliability', in N.J. Salkind (ed.), *Encyclopedia of research design*, Vol. 3, pp. 1495–1498, viewed 15 April 2019, from https://books.google.co.za/books?hl=en&lr=&id=pvo1SauGirsC&oi=fnd&pg=PA1149&dq=Encyclopedia+of+Research+Design,+Volume+2+edited+by+Neil+J.+Salkind&ots=qtbOzt-g5&sig=tXKbtfcQjNOa0qRT1tcmQInrok4&redir_esc=y#v=onepage&q=test-retest%20&f=false.
- Pisani, L., Borisova, I. & Dowd, A.J., 2015, *International development and early learning assessment technical working paper*, viewed 20 May 2019, from http://resourcecentre.savethechildren.se/sites/default/files/documents/idela_technical_working_paper_v3_nodraft.pdf.
- Pisani, L., Dyenka, K., Sharma, P., Chhetri, N., Dang, S., Gayleg, K. et al., 2017, 'Bhutan's national ECCD impact evaluation: Local, national, and global perspectives', *Early Child Development and Care* 187(10), 1511–1527. <https://doi.org/10.1080/03004430.2017.1302944>
- Polit, D.F., 2014, 'Getting serious about test-retest reliability: A critique of retest research and some recommendations', *Quality of Life Research* 23(6), 1713–1720. <https://doi.org/10.1007/s11136-014-0632-9>
- Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Ip, P. et al., 2014, *Report on validation, finalization and adoption of the East Asia-Pacific early child development scales (EAP-ECCDS)*, UNICEF, viewed 11 May 2021, from <https://arnc.net/eccd-scales-detail?id=24>.
- Rust, J. & Golombok, S., 2014, 'Reliability', in *Modern psychometrics: The science of psychological assessment*, pp. 72–78, viewed 10 July 2019, from <https://books.google.co.za/books?id=Vu8ABAAQBAJ&pg=PA71&dq=retest+reliability+standardisation&hl=en&sa=X&ved=0ahUKewitr4mz3frhAhVjs3EKhcQTCsYQ6AEINjAD#v=onepage&q=retest%20reliability%20standardisation&f=false>.
- Salthouse, T.A., 2005, 'Relations between cognitive abilities and measures of executive functioning', *Neuropsychology* 19(4), 532–545. <https://doi.org/10.1037/0894-4105.19.4.532>
- Snelling, M., Dawes, A., Biersteker, L., Girdwood, E. & Tredoux, C., 2019, 'The development of a South African early learning outcomes measure (ELOM): A South African instrument for measuring early learning program outcomes', *Child: Care, Health and Development* 45(2), 257–270. <https://doi.org/10.1111/cch.12641>
- Soares, M.A. & McCrimmon, A.W., 2013, 'Test review: Wechsler preschool and primary scale of intelligence – Fourth edition: Canadian', *Canadian Journal of School Psychology* 28(4), 354–351. <https://doi.org/10.1177/0829573513497343>
- Stein, D.J., Koen, N., Donald, K.A., Adnams, C.M., Koopowitz, S., Lund, C. et al., 2015, 'Investigating the psychosocial determinants of child health in Africa: The Drakenstein child health study', *Journal of Neuroscience Methods* 252, 27–35. <https://doi.org/10.1016/j.jneumeth.2015.03.016>
- Swank, J. & Mullen, P., 2017, 'Evaluating evidence for conceptually related constructs using bivariate correlations', *Measurement and Evaluation in Counseling and Development* 50(4), 270–274. <https://doi.org/10.1080/07481756.2017.1339562>
- Syeda, M.M. & Climie, E.A., 2014, 'Test review: Wechsler preschool and primary scale of intelligence – Fourth edition', *Journal of Psychoeducational Assessment* 32(3), 265–272. <https://doi.org/10.1177/0734282913508620>
- Thorndike, T., 2014, 'Test review of Wechsler preschool and primary scale of intelligence – Fourth edition', in J.F. Carlson, K.F. Geisinger & J.L. Jonsen (eds.), *The nineteenth mental measurements yearbook*, pp. 15–19, Buros Center for Testing, Lincoln Nebraska, NE.
- United Nations, n.d., *Sustainable development goals: 4 quality education*, viewed 10 April 2019, from <https://www.un.org/sustainabledevelopment/education/#tab-bec3d6b2e412d024e05>.
- Warner, R.M., 2013, 'Reliability, validity, and multiple-item scales', in *Applied statistics: From bivariate through multivariate techniques*, pp. 901–952, viewed 10 August 2019, from <https://books.google.co.za/books?id=b1bXhepuJOEC&pg=PT940&dq=pearson%27s+r+test-retest&hl=en&sa=X&ved=0ahUKewjK-LewuvjhAhVHQrUIHedABNs4ChDoAQg6MAM#v=onepage&q=pearson's%20r%20test-retest&f=false>.
- Wechsler, D., 2012a, *WPPSI-IV: Wechsler preschool and primary scale of intelligence*, NCS Pearson, Inc., San Antonio, TX.
- Wechsler, D., 2012b, *WPPSI-IV technical and interpretive manual*, NCS Pearson, Inc., San Antonio, TX.