



The construction and evaluation of an English Exam Skills test for primary school students with dyslexia

Edmen Leong^{1*} and Hu Guangwei²

1 Dyslexia Association of Singapore

2 The Hong Kong Polytechnic University

Abstract

In this article, the construction of a test previously used successfully with dyslexic children and low achievers to assess performance is formally evaluated. An English Exam Skills Programme (EESP) was developed and implemented in 2013 with the goal of helping primary school students with dyslexia develop their English Language skills and achieve in their school and national examinations. The design of the EESP adhered to the Orton Gillingham principles and aimed to ensure that the pedagogy would allow students to transfer skills and concepts learnt to their examination performance. Leong (2015) reports a study conducted to evaluate the progress of students in the EESP using a pre-test and a post-test design. Results from the study suggested that the EESP was effective in addressing the English Language development and examination needs of primary school learners with dyslexia. The study however was based on tests that were designed by the curriculum developers of the EESP and were not subjected to a full validation process. To accurately establish the effectiveness of the EESP, it is also important to ensure that the testing procedures used are optimally reliable and valid. In order to achieve this goal, a new English Exam Skills test for primary students enrolled in the EESP has been developed and validated, following McNamara's (2000) "testing cycle" of the design stage, the construction stage, the try-out stage, and the operational stage. Results obtained from the trialling and validation of the test, including item and whole test analyses, were used to refine and finalize the test. Test takers' performances on this test (both the original and revised versions) were compared with their performances on a test conducted in mainstream primary schools. Substantial correlations constituted evidence of convergent validity. The analyses not only helped to establish the construct validity of the newly developed test but made it possible to predict EESP students' performance on their school and national examinations. In addition, such analyses helped to gauge the effectiveness of the EESP curriculum and the English Language development of dyslexic learners in primary schools.

Keywords: Dyslexia, English Exam Skills, Language testing and assessment, Test validation, Test design

* Correspondence to: Edmen Leong, Dyslexia Association of Singapore. Email: edmen@das.org.sg

INTRODUCTION

The English Exam Skills Programme (EESP) was established in 2013 by a group of educational therapists in the Dyslexia Association of Singapore (DAS) with the goal of helping their primary school students with dyslexia achieve in their school and national examinations. Special effort was made in the development of the EESP curriculum to adhere to the Orton Gillingham principles so that lessons would be delivered in a direct, explicit, progressive, and multisensory manner (Leong, Asjamiah, & Wang, 2017) and students would be able to transfer concepts and skills learnt in EESP to answering questions in their examination papers. The EESP curriculum developers used curriculum design processes adapted from Nation and Macalister (2010) as well as Richards (2001) as guidelines to ensure that the curriculum designed and implemented would be in line with the goals of the EESP.

Leong (2015) reports a study conducted to evaluate the progress of students in the EESP. In this study, a pre-test and a post-test were administered to evaluate the effectiveness of the Programme. A quantitative comparison of a synthesis and a transformation component of the pre-test and the post-test revealed the progress made by students who completed the Programme. Results from this study suggested that the EESP was effective in addressing the English examination needs of primary school learners with dyslexia. A subsequent study conducted by Leong, Asjamiah, and Wang (2017) explored classroom practices in the EESP classroom. The study found that the progress of enrolled students was attributable to how the educational therapists taught the classes. The systematic, progressive, and multisensory teaching practices contributed to the students' success in the Programme. The most recent study on this approach, by Elfira, See, Tan, and Leong (2018), found that the designed Programme had similar benefits for struggling learners who were not diagnosed with dyslexia.

While the findings from the above studies on the EESP were encouraging, it is important to note that they were based on tests designed by the curriculum development team that had not been subjected to a full validation process. To firmly and accurately establish the effectiveness of the EESP, it is also important to ensure that the testing procedures used to evaluate this effectiveness are optimally reliable and valid. In order to achieve this goal, a new English Exam Skills test for Primary three and four students enrolled in the EESP has been developed and validated, following McNamara's (2000) "testing cycle" of the design stage, the construction stage, the try-out stage, and the operational stage. Results obtained from the trialling and validation of the test, including item and whole test analyses, were used to evaluate and refine the test. The performance of test takers on this test was correlated with their performance on a Primary 4 English language test modelled on the PSLE exam and constructed from a mixture of items from school test papers set by various schools. The item and whole test analyses not only helped to establish the construct validity of the newly developed test, but also made it possible to predict EESP students' performance on their school and national

examinations. In addition, such analyses helped to further gauge the effectiveness of the EESP curriculum by ensuring robust testing procedures.

RESEARCH AIMS

Given the importance of having a valid and reliable measure of the progress of the EESP students, the primary aim of this study is to construct an achievement test for primary three and four children enrolled in the EESP that has desirable psychometric properties, namely reliability and validity. While it is imperative that the designed test accurately measures the achievements of EESP students as they progress academically, it is also important that the test can reasonably well predict the students' progress as evaluated by their mainstream school tests. The secondary aim of this study is therefore to compare the designed test with a test of similar language knowledge and skills that would be typically administered in Singapore primary schools. Finally, it is also desirable to track the progress of these students undergoing the EESP with the new test.

In summary, the primary and secondary research aims of this study are as follow:

1. to construct and validate a new achievement test for the EESP;
2. to compare how target students fare on the new test and a test of similar language knowledge and skills typically administered in Singapore primary schools;
3. to ascertain if the new test is able to track the progress of students in the EESP.

METHOD

Research design

To address the primary and secondary research aims, two tests were prepared for this study: a self-developed EESP test (to be described in the next section) and a school-based test mirroring a typical test administered in Singapore primary schools. The school-based test was constructed by selecting sections from past English exam papers set by various primary schools. This was to ensure that none of the students participating in the study would have taken the test previously. This school-based test was then distributed to five primary school teachers, who helped with ensuring that the contents of the test were in line with a typical test in primary schools. The research design is described in stages below and summarised in Figure 1.

Stage 1 (Week1):

Students who enrolled for the EESP Programme sat for both the designed EESP test and the school-based test upon enrolment.

Stage 2 (Week 2 to week 39):

Upon completion of the tests, the students underwent 3 terms to a year of the EESP intervention. The EESP test as a whole and its individual items were analysed, and problematic test items identified in the process were removed from the test. These procedures were intended to achieve the first research aim of this study, that is, to construct and validate a new EESP test. Items on the school-based test were rearranged to create an alternate form of the test for use as a post-test.

Stage 3 (Week 40):

Upon completion of the EESP intervention, students sat for the revised EESP test and the alternate form of the school-based test.

Stage 4 (After week 40):

The students' performances on the EESP test (both the original and the revised version) and the school-based test (both the pre-test and the post-test) were analysed to achieve the two secondary aims of this study. First, a Pearson's correlation analysis was conducted to assess the convergence of the EESP test and the school-based test. Second, paired-samples t-tests were run to compare the pre-test and post-test scores of both the EESP test and the school-based test. In all these analyses, the EESP test items that were found problematic and dropped from the revised version were removed from the computation of the pre-test and post-test scores.

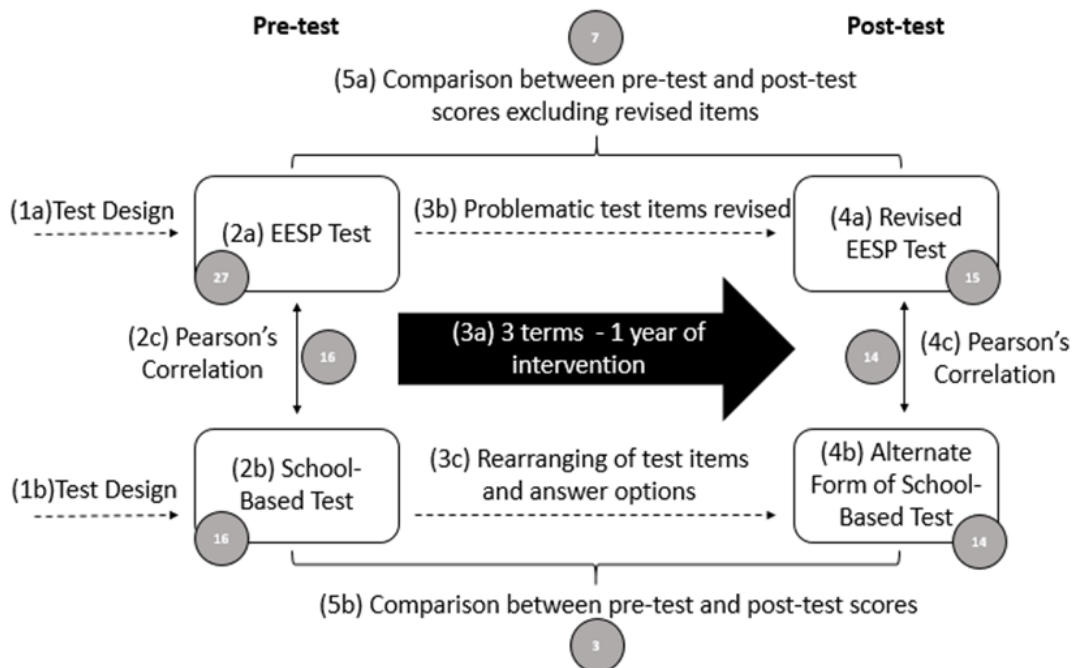


Figure 1. Research design: tests, procedures, and analyses

Participants

Primary 3 and 4 students diagnosed with dyslexia and enrolled in the EESP participated in the study. Since students enrolled and left the EESP throughout the 4-term duration of the study, and because there were absentees during some of the test sessions, the number of participants taking each test was different. A total of 27 students took the EESP pre-test, 15 students sat for the EESP post-test, and 7 students participated in both. A total of 16 students took the school-based pre-test, 14 students sat for the school-based post-test, and 3 students participated in both. The EESP test and item analyses were based on the data from the 27 students who took the EESP pre-test. The comparison between EESP pre- and post-test performances, however, was based on the data from the 7 students who participated in both tests and went through the intervention Programme. The comparison between the school-based pre- and post-test performances was also based on the 3 students who participated in both the pre- and post-tests and went through the intervention. Table 1. Summarizes the mean age and gender of the students who were part of the item analysis, and EESP and school-based pre- and post-test analyses.

Table 1. Summary of profile of students who participated

	No. of students	Gender		Mean age
Item Analyses	27	12 boys	5 girls	9.81
EESP Pre- and Post-test	7	6 boys	1 girl	9.57
School-based Pre- and Post-test	3	3 boys	1 girl	10

EESP Test content and structure

The EESP test was designed as an achievement test to assess mastery of curricular content covered in the Programme. The curricular content includes grammar, editing, and reading comprehension. These curricular areas have been selected to be part of the Programme since the EESP students are particularly weak in them. Grammar is also an area that students need to master sufficiently in order to answer several types of questions commonly found in their school exam papers. The newly designed EESP test had 31 questions allotted a total of 37 marks, including 12 one-mark grammar questions (GR1 – GR12), 11 one-mark editing questions (ED1 – ED11), and 8 comprehension questions assessing 14 sub-skills and attracting 14 marks (CP1 – CP14). These questions tested for the grammatical concepts, editing abilities, and comprehension skills taught and developed in the EESP throughout the year of intervention. A copy of the test is presented in Appendix A.

Analyses conducted to validate the EESP test

Given the small number of test takers involved, classical item analysis was conducted (Alderson, Clapham, & Wall, 1995). The Facility Value (FV) and the Discrimination Index (DI) of each of the test items were calculated to measure how difficult each test item was and how well an item discriminated between the weak and strong test takers respectively. The FV (ranging from 0 to 1) of each item in the test was computed as the percentage of students who answered the question correctly. An FV of 0 would mean that none of the students answered the question correctly, while an FV of 1 would mean that all students provided the correct answer. Thus, the lower an FV was, the more difficult the test item was. Following Alderson et al., (1995), the 27 students who took the pre-test were divided equally into 3 groups according to their total test scores to calculate the DI values. Nine students with the highest scores were classified as the Top Group, the 9 students with the lowest scores as the Bottom Group, and the remaining 9 as the Middle Group. The DI (ranging from -1 to +1) of each item was calculated with the following formula:

$$DI = (RT - RB) / NT$$

Where RT and RB refer to the number of students in the Top Group and the Bottom Group who correctly answered the question respectively, and NT is the number of students in the Top or Bottom Group. A positive DI value would mean that more top than bottom students answered the question correctly, and a negative DI value would mean that more bottom than top students gave the correct answer. The greater the DI value was, the better the test item discriminated between high- and low-performing students. A multiple-choice question (MCQ) analysis was also conducted to evaluate how well the options provided for each question worked. The analysis involved examining the spread of answer choices of the students in the Top, Middle, and Bottom Groups. The analyses described above were intended to identify good and problematic test items. Based on the results of the FV, DI and MCQ analyses, the well-performing test items were kept intact in the revised version of the EESP test, and the problematic items were revised for validation in a new testing cycle.

Whole test analyses were also conducted to assess how reliable the revised EESP test was and to obtain different kinds of validity evidence (Bachman & Palmer, 1996; Popham, 2003). Following the formula and procedure presented by Alderson et al., (1995, p.280), split half reliability estimates were calculated for the revised EESP (both the pre-test and the post-test). Pearson's correlational analyses were run between the EESP test and the school-based test administered as the pre-tests and post-tests to obtain evidence of convergent validity, one manifestation of construct validity. Convergent validity measures how much tests that are theoretically related are actually related, whereas construct validity reflects the degree to which a test measures what it purports to measure. Paired-samples t-tests were also conducted as described in Stage 4 to gather evidence of the EESP test's sensitivity to test takers' progress, which could be

taken as a reflection of the test's content validity (how well it measures the different facets of learning). If the test takers truly made progress as a result of participating in the EESP Programme, they could be reasonably expected to gain scores from the pre-test to the post-test, if the test covered the curricular content adequately (Popham, 2003).

FINDINGS AND DISCUSSION

Results of the EESP Test and Item Analyses

To evaluate the EESP Test, the following steps were taken. The test items and test scores of the 27 students who took the EESP test were analysed. Their test scores are presented in Appendix B. The descriptive statistics for the whole test are presented in Table 2. The mean score indicated that the students answered, on average, a minority of the test questions correctly. Both the large standard deviation and the wide range suggested that there was considerable variation among the students.

Table 2. Descriptive statistics for the EESP pre-test

Maximum score possible	37
Mean	15.60
Standard deviation	4.98
Standard error	0.96
Highest score	26
Lowest score	7
Range	19

The FVs and DIs for the test items are presented in Table 3. Items with FVs close to 0 or 1 and DIs below 0.3 were investigated. The FVs and DIs indicated that of the 37 test items, 19 items functioned well. These items were able to distinguish the top-performing and the low-performing students. They were also of a moderate difficulty level, apparently appropriate for this specific group of learners with dyslexia. These 19 questions were therefore retained in the revised EESP test. The remaining items of the test were investigated closely in order to ensure that an appropriate range of difficulty was maintained to allow test takers to demonstrate improvement from pre to post-test. GR5 and CP9 had both low FVs and negative DIs. Under normal circumstances, such items should be removed from a norm-referenced test (Alderson et al., 1995). However, items with low FVs have a place in an achievement tests such as the present one if few of the test takers have mastered the curricular content (Davidson, 2000; Hughes, 2003). Indeed, this was the case with GR5 and CP9 because a great majority of the students did not

Table 3. Results of item analyses, decisions, and justifications

ITEM	FV	DI	POTENTIAL PROBLEMS	DECISION	JUSTIFICATION FOR KEEPING
GR1	0.48	0.33		Retained	
GR2	0.37	0.22	Low DI	Revised	
GR3	0.48	0.67		Retained	
GR4	0.26	0.56		Retained	
GR5	0.26	-0.11	Negative DI	Retained	Only a minority of students had the required knowledge. There was some guesswork.
GR6	0.67	0.22	Low DI	Revised	
GR7	0.37	-0.22	Negative DI	Revised	
GR8	0.41	0.33		Retained	
GR9	0.59	0.22	Low DI	Revised	
GR10	0.74	0.33		Retained	
GR11	0.70	0.11	Low DI	Revised	
GR12	0.30	0.00	Low DI	Retained	A great majority of students lacked the relevant knowledge.
ED1	0.96	0.11	Low DI	Revised	
ED2	0.37	0.56		Retained	
ED3	0.41	0.67		Retained	
ED4	0.41	0.78		Retained	
ED5	0.74	0.56		Retained	
ED6	0.11	0.33		Retained	
ED7	0.00	0.00	Low FV and DI	Revised	
ED8	0.78	0.44		Retained	
ED9	0.19	0.22	Low FV and DI	Revised	
ED10	0.67	0.33		Retained	
ED11	0.85	0.22	Low DI	Revised	
CP1	0.78	0.22	Low DI	Retained	A low DI was acceptable for an easy item in an achievement test.
CP2	0.44	0.67		Retained	
CP3	0.00	0.00	Low FV and DI	Retained	No student had the required skill, which would be covered in the EESP Programme.
CP4	0.19	0.44		Retained	
CP5	0.33	0.33		Retained	
CP6	0.19	0.22	Low DI	Retained	Few students knew the tested vocabulary to be covered in the EESP Programme.
CP7	0.04	0.11	Low FV and DI	Retained	Few students knew the tested vocabulary to be covered in the EESP Programme.
CP8	0.04	0.11	Low FV and DI	Retained	Few students knew the tested vocabulary to be covered in the EESP Programme.
CP9	0.04	-0.11	Low FV and negative DI	Retained	All students lacked the target skill of inferring. The only correct answer was guesswork.
CP10	0.67	0.33		Retained	
CP11	0.30	0.33		Retained	
CP12	0.22	0.22	Low DI	Retained	A great majority of the students lacked knowledge of what an adverb is.
CP13	0.74	0.33		Retained	
CP14	0.52	0.67		Retained	

The MCQ analysis mentioned above was conducted on the options provided for each multiple-choice question to evaluate how well the correct answer and distractors worked, following Alderson et al's, (1995) approach. This analysis yielded further insights into why certain test items did not function well. For example, the grammar question below (GR2) had a low DI, and the MCQ analysis suggested that option 2 was a wasted distractor because none of the students selected it (see Table 4). Since the present participle (-ing) of verbs would normally co-occur with an auxiliary verb, option 2 was plainly incorrect to all students who took the test. Furthermore, option 4 obviously distracted the wrong students (i.e., 3 students in the Top Group but none in the Bottom Group) due to the ambiguity of the tenses involved (past or present). The rest of the GR questions were analysed similarly, and the results are presented in Appendix C.

2. This basketball team _____ three players who are over six feet tall.
- 1) have 2) having
- 3) has 4) had

	1	2	3*	4	Total
Top Group	2	0	4	3	
Middle Group	3	0	4	2	
Bottom Group	7	0	2	0	
Total	12	0	10	5	27

Based on the detailed item analyses presented above, a revised version of the EESP test was produced. Of the 37 items in the original EESP test, 28 were identified as functioning well and were retained in the revised EESP test, and the remaining 9 were revised for further work. These revised items were excluded from the further analyses conducted for this study. The 28 items kept unchanged in the revised EESP test were administered as the post-test, as described in Stage 4a of Figure 1. The descriptive statistics for the post-test scores are presented in Table 5. As the pre-test contained more items than the post-test and because different students took the pre-test and post-test (with the exception of the 7 students who took both tests), no direct comparison can be made between the descriptive statistics presented in Table 2 and Table 5. However, it is still important to note that the mean score for the post-test was considerably higher than that for the pre-test relative to the maximum score possible, and there was less heterogeneity, as indexed by the standard deviation and the range, in the post-test performances than in the pre-test ones.

Table 5. Descriptive statistics for the EESP post-test

Maximum score possible	28
Mean	13.36
Standard deviation	4.63
Standard error	1.24
Highest score	21
Lowest score	6
Range	15

The revised 28-item EESP test was divided into an odd-numbered and an even-numbered group for the calculation of its split half reliability indices (Alderson et al., 1995). The split half reliability, obtained with the Spearman Brown correction formula, was .80 for the pre-test scores (obtained at Stage 2a in Figure 1) and .78 for the post-test scores (obtained at Stage 4a in Figure 1). Given the relatively small numbers of test takers involved and test items included, these reliability indices were acceptable for an achievement test (Hughes, 2003).

Convergent validity

As described earlier, Pearson's correlational analyses were conducted on the scores of the revised EESP test and the school-based test to gauge convergent validity, a subtype of construct validity. Given that the two tests assessed overlapping constructs of English

language achievement, the scores of the students taking the two tests concurrently would be significantly correlated if there was good convergent reliability. The correlation between the EESP pre-test and the school-based pre-test was substantive, $r(14) = .577$, $p = .019$. This was an acceptable index of convergent validity since the two tests did not measure exactly the same knowledge/skills, although there was considerable overlap. Indeed, when the items of the school-based test unrelated to the 3 areas (i.e., grammar, editing, and reading comprehension) assessed by the EESP test were excluded from the correlational analysis, an exceptionally strong correlation was found, $r(14) = .946$, $p < .001$. The moderate relationship indexed by the first coefficient suggested that each test assessed some knowledge and skills unique to it, and the strong relationship indexed by the second coefficient indicated that students' performances for similar components (i.e., grammar, editing, and reading comprehension) of the EESP test and the school-based test were highly comparable. Thus, students' performances on the revised EESP test could be used to predict their performances on a mainstream exam paper such as the school-based test used in this study.

Similar correlational analyses were run on the students' scores for the EESP post-test and the school-based post-test. The correlation between the revised 28-question EESP test and the full school-based test was not statistically significant, $r(12) = .449$, $p = .107$. A similar correlation coefficient, $r(12) = .491$, $p = .075$, was obtained when only the relevant areas (i.e., grammar, editing, and reading comprehension) of the school-based post-test were included in the analysis. The two correlation coefficients were not statistically significant largely due to the small sample size, nevertheless, their magnitudes still indicated medium effect sizes based on the benchmarks recommended by Plonsky and Oswald (2014) for language learning studies. In comparison with the correlations observed in the pre-tests, the reduced strength of the correlations between the post-tests appeared to be a function of the effectiveness of the EESP Programme (see the results of the analyses reported in the following section). Thanks to the EESP Programme, the students made significant gains from the EESP pre-test to the EESP post-test, and the inter-student variability in test performance was reduced. On the other hand, the students did not make equal improvements from the school-based pre-test to the school-based post-test, some of whose content was not covered by the EESP Programme. These factors worked together to reduce the strength of the correlations found for the post-test data.

Taken together, the results reported above provided consistent and robust evidence of the revised EESP test's convergent validity and, by extension, construct validity. They also indicated that the revised EESP test contributed some unique variance to the construct of English language achievement over and beyond the school-based test.

Sensitivity to student progress and content validity

As explained earlier, when the students enrolled in the EESP Programme mastered the curricular content well, they would be able to make significant gains from the pre-test to

the post-test if the test sampled the curricular content adequately. In other words, if the revised EESP test had good content validity, it should be able to track the students' progress if they indeed improved in their relevant language knowledge and skills as a result of participating in the EESP Programme. Therefore, a paired-samples t-test was run to compare the pre-test and post-test scores of the 7 students who took both versions of the EESP test. For the sake of comparison, another paired-samples t-test was conducted to determine if the 3 students who took both the school-based pre-test and post-test improved significantly in their scores over time. The results of the t-tests are summarized in Table 6. The mean scores for the EESP pre-test and post-test showed drastic improvements, and the difference was statistically significant at $p < .001$. The effect size, as measured by Cohen's d , was a large one by conventional standards (Cohen, 1988). Notably, the standard deviation for the post-test scores was markedly smaller than that for the pre-test scores, indicating that the 7 students performed more similarly to each other on the post-test than on the pre-test. Both the statistically significant difference between the mean scores and the markedly smaller standard deviation for the post-test scores constituted clear evidence that the EESP Programme was effective in improving the students' language knowledge and skills assessed by the revised EESP test. These results were consistent with the findings of previous studies conducted on the progress of students in the EESP Programme (Leong, 2015; Leong et. al, 2017; Elfira et. al. 2018).

Table 6. Results of t-tests on the pre- and post-test of the revised EESP and the school-based test

Measure	Pre-test		Post-test		t	df	p	d
	M	SD	M	SD				
EESP test	7.71	4.61	14.60	3.95	-6.79	6	.000	1.6
School-based test	5.33	4.04	10.33	3.06	-3.27	2	.041	1.4

As only 3 students took both the school-based pre-test and post-test, the t-test results obtained with such a small sample size were only indicative. The results of the t-test also revealed a statistically significant improvement from the pre-test to the post-test. The effect size measured by Cohen's d was large in itself but smaller than that observed for the revised EESP test. Taken together, these tentative results suggested that the students made relatively less improvement in language knowledge and skills that were not covered by the EESP Programme but were assessed in the school-based test. This differential pattern of improvement on the revised EESP test and the school-based test contributed to the lower correlation found between the post-tests reported in the preceding section.

CONCLUSION

This study has provided clear evidence of reliability and validity for a newly developed language test for students with dyslexia. The classical item analyses and whole test analyses (chosen over the more sophisticated statistical apparatus for test validation because of the small sample sizes available) have revealed that the revised EESP test has desirable psychometric properties to function as an assessment tool for research and educational purposes. Equally importantly, the study has provided a robust framework for developing and validating a new test for use in educational contexts such as ours. In such contexts, students who are available to participate in the validation activities are often rather limited in number, making it impossible to employ more advanced analyses such as Rasch modelling (Chen, Lenderking, Jin, Wyrwich, Gelhorn, & Revicki, 2014). As this study has demonstrated, such restrictions do not mean that we should give up the goal of ensuring the validity and reliability of a new test.

The process of analysing individual test items, as we have done in our study, allows the test developers and curriculum designers additional insights into the strengths, weaknesses, and thought processes of students who completed the test. Thus, engagement in such analyses could lead the test and curriculum developers to better understand the educational needs of the target students and to make further enhancements to the educational Programmes developed for them. In our future research, the validated EESP test will be used with new cohorts of students enrolled in the EESP Programme to determine if the effectiveness of the EESP Programme can be replicated with a view to further improving this programme. Meanwhile, we also plan to start a new testing cycle to validate newly designed items as well as the items culled from the EESP test and revised for further research. Only when a sufficiently large pool of good test items is created can we construct parallel versions of the EESP test that are needed for further research on whether and how explicit and progressive teaching can benefit learners with dyslexia.

REFERENCES

- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.
- Chen, W-H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, 23, 485-493.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Davidson, F. (2000). The language tester's statistical toolbox. *System*, 28, 605-617.
- Elfira, T., See, E., Tan, S. H. J., & Leong, E. (2018). Exploring the effectiveness of the English Examination Skills Programme on struggling non-dyslexic learners. *Asia Pacific Journal of Developmental Differences*, 5(2), 141-162.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Leong, E. (2015). Improving English exam skills for dyslexics in primary education in Singapore. *Asia Pacific Journal of Developmental Differences*, 2(2), 184-201.
- Leong, E., Asjamiah, S., & Wang, A. (2017). Exploring the classroom practices of the English Exam Skills Programme for Singaporean primary school children. *Asia Pacific Journal of Developmental Differences*, 4(2), 167-194.
- McNamara, T. (2000). *Language testing*. Oxford, UK: Oxford University Press.
- Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. London, UK: Routledge.
- Plonsky, L., & Oswald, L. F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878-912.
- Popham, W. J. (2003). *Test better, teach better: The instructional role of assessment*. Alexandria, VA: Association for Supervision and Curriculum Development
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge, UK: Cambridge University Press.

APPENDIX A

6. These are _____ presents. The rest are _____.
1) you're.....there's 2) your's.....theirs
3) yours.....their 4) your.....theirs ()
7. Neither John nor his classmates _____ their homework.
1) has done 2) has did
3) have done 4) have did ()
8. The twins _____ home when they found a wallet on the pavement.
1) walk 2) was walking
3) were walking 4) walked ()
9. If she _____ early, who would not have woken up late.
1) has slept 2) had slept
3) have slept 4) had slept ()
10. Lisa _____ home early to catch her flight to Japan tonight.
1) is leaving 2) was leaving
3) were leaving 4) was left ()
11. "It is alright! We can do this _____!" the students said to their teacher.
1) ourself 2) yourself
3) ourselves 4) ourselve ()
12. The class of students _____ on a field trip to Malacca this weekend.
1) are going 2) was going
3) were going 4) is going ()

APPENDIX A**Part 2: Editing**

Fill in the boxes with the correct spelling of the words in bold. (11 Marks)

Susan was **swiming** under water when something came near her. It had a round

middal with eight long arms **stiking** out of it. Each arm had rows of little cups that could

hold things tightly. In its **sentre** was a mouth with two beaks as **shurp** as knife blades.

In the past, many **sailers** had been afraid of this animal. But Susan was not afraid.

She was a scientist learning about the octopus. She knew that it eats crabs and other

small shellfish from the bottom of the sea. She knew that many **octopuss** give off lots of

ink when they feel in **danjer**. This makes a **screan** so their enemies can't see them slip

away. Susan **hopped** this would not happen because she wanted to **wach** the octopus

so she could learn more about it.

Adapted from early reading comprehension in varied subject matter – Jane Ervin

APPENDIX A

Part 3: Comprehension (Open-Ended)

Read the following passage.

A deer went for a drink and saw his reflection in the water. "What magnificent antlers I have!" he exclaimed proudly. Then, he looked down at his legs and sighed.

"What's wrong?" asked a fish.

"I hate my long and thin legs. I would rather have no legs like you," 5
groaned the deer.

"I doubt it," replied the fish, but before it could say more, the forest was filled with the barking of hunting dogs.

That prompted the deer to take to his heels and flee like the wind to safety. Just as he thought he was safe, his antlers got caught in the branches of a tree. 10
They caused a nest to fall off the tree.

"Don't struggle," said a bird, "or you will become more tangled. The dogs will be here and those antlers will be the death of you."

Though the deer's every instinct was to struggle, he took the bird's advice. He managed to free himself calmly soon after and return safely to his family. 15

The deer heaved a sigh of relief and said, "I've just realised something. The part of me I was most proud of almost killed me and the part I despised most has saved me."

APPENDIX A

Based on the above passage, answer the questions below.

1. Write down one word in line 1 that has the same meaning as an 'image.' (1m)

2. Box up the adjective in this sentence - "What magnificent antlers I have!" (1m)

"What magnificent antlers I have!"

3. What does 'That' in line 9 refer to? (1m)

4. What does 'They' in line 11 refer to? (1m)

- 5a Read the question below.

Circle the question word. (1m)

Underline the verb. (1m)

Box up the target words. (1m)

Bracket the quoted words. (1m)

Which four-word phrase in paragraph 5 has the same meaning as 'ran'?

- 5b Write down your answer for the above question in the line below. (1m)

APPENDIX A

6a Put a tick in the box next to the sentence that is correct. (1m)

An adverb tells you how something look like.	
An adverb tells you how an action is done.	

6b Box up the adverb(s) in the sentence below. (2m)

He managed to free himself calmly and return safely to his family.

7. Which part of its body did the deer not like? (1m)

8. Which word in paragraph 8 has the same meaning as the word 'hated'? (1m)

Component	Grammar (GR)									Editing (ED)								Comprehension (CP)																Total	Cat.
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	13	14									
Item	0	0	0	1	0	1	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	7								
Student 2	1	0	1	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	9								
Student 3	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	9								
Student 4	0	0	0	1	1	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	9								
Student 5	0	1	1	0	0	1	1	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	11								
Student 6	0	0	0	0	1	0	1	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	11								
Student 7	1	1	0	0	0	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	11								
Student 8	0	0	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	1	1	0	1	13								
Student 9	1	0	0	0	1	1	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1	13								
Student 10	1	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	0	13								
Student 11	0	1	0	1	0	1	0	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	15								
Student 12	1	0	0	1	0	0	1	1	0	1	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	1	15								
Student 13	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	15								
Student 14	1	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0	1	0	0	0	1	0	0	1	1	15								
Student 15	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	1	15								
Student 16	0	1	0	0	1	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1	1	16								
Student 17	0	1	0	0	0	1	0	0	1	1	0	1	0	0	1	0	0	1	0	0	1	1	0	0	1	1	17								
Student 18	1	1	1	0	0	0	1	1	1	1	0	1	0	0	1	0	0	0	0	0	0	0	1	0	1	0	17								
Student 19	1	1	1	0	0	0	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	1	0	1	1	17								
Student 20	1	0	1	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0	1	1	19								
Student 21	0	1	1	1	0	1	0	1	1	1	0	1	1	0	0	1	0	0	1	1	0	0	0	0	0	0	19								
Student 22	1	0	1	1	1	0	0	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	20								
Student																																			

APPENDIX C—MCQ ANALYSIS

MCQ distractor analysis for item GR2					
	1	2	3*	4	Tot.
HA	2	0	4	3	
MA	3	0	4	2	
LA	7	0	2	0	
Tot.	12	0	10	5	27

MCQ distractor analysis for item GR5					
	1	2	3*	4	Tot.
HA	3	5	1	0	
MA	0	7	0	2	
LA	4	4	1	0	
Tot.	7	16	2	2	27

MCQ distractor analysis for item GR6					
	1	2	3*	4	Tot.
HA	1	0	1	7	
MA	1	1	1	6	
LA	3	0	1	5	
Tot.	5	1	3	18	27

MCQ distractor analysis for item GR7					
	1	2	3*	4	Tot.
HA	5	1	2	1	
MA	5	0	4	0	
LA	4	0	4	1	
Tot.	14	1	10	2	27

MCQ distractor analysis for item GR9					
	1	2*	3	4	Tot.
HA	1	2	6	1	
MA	3	3	2	0	
LA	3	2	3	1	
Tot.	7	7	11	2	27

MCQ distractor analysis for item G11					
	1	2	3*	4	Tot.
HA	1	0	7	1	
MA	2	0	6	1	
LA	2	1	6	0	
Tot.	5	1	19	2	27

MCQ distractor analysis for item GR12					
	1	2	3*	4	Tot.
HA	5	1	0	3	
MA	2	1	4	2	
LA	5	1	0	3	
Tot.	12	3	4	8	27