

# *Dynamic tests as a language-free method for assessing reading in a multilingual setting*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Pye, R. E. ORCID: <https://orcid.org/0000-0002-1395-7763> and Chan, H. H. (2023) Dynamic tests as a language-free method for assessing reading in a multilingual setting. *Journal of Cultural Cognitive Science*, 7 (2). pp. 147-158. ISSN 2520-1018 doi: 10.1007/s41809-023-00120-8 Available at <https://centaur.reading.ac.uk/111330/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1007/s41809-023-00120-8>

Publisher: Springer Nature

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



RESEARCH PAPER

# Dynamic tests as a language-free method for assessing reading in a multilingual setting

Rachel E. Pye · Huey Hwa Chan

Received: 30 September 2022 / Revised: 10 February 2023 / Accepted: 4 March 2023  
© The Author(s) 2023

**Abstract** It is difficult to assess reading ability or diagnose dyslexia in Malaysia due to its complex linguistic background of simultaneous multilingualism and multiliteracy. Low scores on traditional reading tests could be explained by language and contextual factors rather than indicating reading difficulty. One method shown to be less sensitive to specific language factors is the Dynamic Testing paradigm (Elbro et al., 2012). Fifty-nine Malaysian adult readers of English and Bahasa Melayu and/or Chinese were recruited and their performances on the dynamic test assessed. Scores on the Dynamic Test significantly predicted potential reading difficulty as measured by the Adult Reading History Questionnaire (Lefly & Pennington, 2000) but not reading proficiency measured by a Lexical Decision Task (LDT). Participants were then grouped by language of first literacy: alphabetic (English and Bahasa Melayu) or morphosyllabic (Chinese). No significant differences were found between those reading morphosyllabic versus alphabetic orthographies for the Dynamic Test. The Dynamic Test was a significant predictor of potential reading difficulty in the alphabetic group only. Thus, it was concluded that the

dynamic format could be a suitable assessment tool for multilinguals, but that further research is required to determine its utility, especially in non-alphabetic languages.

**Keywords** Dynamic testing · Multilingual · Reading difficulty · Morphosyllabic · Multiliteracy · Alphabetic

## Introduction

Malaysia is a multiethnic and multilingual country in South East Asia. Its national language, Bahasa Melayu (BM), is the language of instruction in public schools and for government functions, and is generally associated with the largest ethnic group, the Malay Malaysians. The second-largest ethnic group is Chinese Malaysians, who speak a variety of dialects such as Hakka, Hokkien, Teochew etc., with Mandarin a language of instruction in the ‘vernacular’ schools which educate approximately 4% of Malaysian school children (Lim, 2017). The third largest ethnic group is Indian Malaysians, and Tamil is also a language of instruction in ‘vernacular’ schools (Department of Statistics Malaysia, 2020). Bahasa Melayu is a required subject in schools using other languages, and English is also a compulsory subject taught in all schools (Gomez, 2004), with every child in Malaysia expected to be functionally bilingual in English and BM by graduation from

R. E. Pye (✉)  
University of Reading, Reading, UK  
e-mail: rachel.pye@reading.ac.uk

H. H. Chan  
University of Reading Malaysia, Johor Bahru, Malaysia

secondary education (Ministry of Education Malaysia, 2013). These languages differ in their orthography: both Bahasa Melayu and English are Latin alphabetic, with standard BM transparent in terms of grapheme-phoneme correspondences, although it is also considered a diglossic language (Jalil & Liow, 2008). Tamil is syllabic and Mandarin is morphosyllabic. Due to the diverse linguistic environment, most Malaysians are at least bilingual and biliterate, though many speak three or more languages. The multilingualism in Malaysia differs from the sequential bi/multilingualism often researched in European contexts, with L1 coming before L2, L3 etc. Instead, these languages often develop simultaneously with children and adults speaking multiple languages within the same family, let alone within communities and institutions.

Research on dyslexia in Malaysia is limited (Gomez, 2004). People with reading disabilities in Malaysia are mainly grouped with others who experience different learning disabilities (Dzalani & Shamsuddin, 2014). The issue of accurate bi/multilingual dyslexia assessment in children and adults is complex (Elbeheri & Everatt, 2016; see Kormos, 2017, for a review). The first question is which language to assess in, or whether to use more than one language, assuming also that these languages are read as well as spoken. Lindgrén and Laine (2007) found that bilingual adults scored significantly lower in standardised tests in their L2 compared to L1. This indicates that there may be an over-diagnosis of reading difficulties in bilinguals which could in fact reflect vocabulary or language differences (Elbro et al., 2012; see Lachmann et al., 2022 for a discussion of discrepancy approaches to dyslexia diagnosis). Elbeheri and Everatt (2016) noted the complexities around using appropriate norms when testing children from nonrepresentative samples; they recommended assessors test in both L1 and L2 when possible, until such time that appropriate measures for that community are created. Kormos's (2017) review recommended diagnostic testing in an L1 wherever possible, but this comes from a perspective of having a minority home language within a dominant community language, which is not appropriate in a context such as Malaysia's. There are currently two assessment batteries in Bahasa Melayu (Lee, 2008; Lee et al., 2020), with the latter battery specifically addressing the issue of norming and

standardising for this multilingual population, albeit only normed for one age group in one region of Malaysia thus far. The use of these batteries for many Malaysian children would therefore be analogous to using a diagnostic test in an L2. To the best of our knowledge, there remain no specific assessment batteries for Malaysian children in English, Mandarin, or Tamil, and no clear approach to determining which of these should be used, should they exist.

One approach to assessing children in diverse linguistic environments is to use a measure that is not contingent upon vocabulary or proficiency in any particular language, but to measure an underlying skill in acquiring reading ability. Testing the potential to learn a particular skill rather than current knowledge/ability is the hallmark of dynamic testing (see Grigorenko & Sternberg, 1998). Elbro et al. (2012) taught symbol-sound pairs (one grapheme to one phoneme) in an artificial orthography, measuring the ability to synthesise these learnt symbol-sound pairs, with the express purpose of identifying possible dyslexia by using a dynamic test approach in an L2 adult sample (Danish learners in Denmark). The dynamic test strongly correlated with traditional word reading measures such as non-word reading and phoneme awareness, but correlated only moderately with what Elbro et al. (2012) called 'environmental factors' such as education and vocabulary. It was sensitive to L1 adults with a diagnosis of dyslexia and L2 adults with suspected dyslexia, leading Elbro et al. (2012) to conclude that both L1 and L2 learners could be tested without reference to their specific language context. Other studies have looked at non-language measures of reading in European bilingual children. Aravena et al., (2013, 2016, 2018) demonstrated group differences in dyslexic and non-dyslexic Dutch children in the learning of artificial symbol-speech sound correspondences across a series of studies. Horbach et al. (2018) showed that a symbol-sound learning paradigm (SSP) predicted reading ability in prereaders three years later, both in monolingual and multilingual German children.

Interestingly, despite Elbro et al.'s (2012) aim to avoid language-specific factors in the identification of adult dyslexics, the authors asserted that "...the dynamic test is limited to alphabetic orthographies... A rather different learning test would be necessary for syllabic or morpho-syllabic orthographies such as the Japanese and Chinese" (page 183). This assertion

reflects recent exploration of the differences and similarities across orthographies related to reading development and dyslexia identification, with the recognition in some theories (e.g. the componential model of reading, Aaron et al., 2008) that environmental factors are important when looking at development of reading, rather than purely cognitive skills. The Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) proposed that first literacy, the language one first learns to read, can impact the strategies used in reading subsequent languages. There is some empirical evidence supporting this (e.g. Chikamatsu, 1996; Wang et al., 2003), perhaps implying a qualitatively different approach to reading across orthographies. However, it should be noted that the Psycholinguistic Grain Size Theory makes the assumption that languages are learnt sequentially, which might not be the case for all bi/multiliterate readers. Daniels and Share (2018) eloquently argued that the over-emphasis of research on European alphabetic languages is problematic and fails to consider the full impact of orthographic complexity on reading, leading to an over-emphasis of the role of phonology in the literature on reading development. However, there is also discussion in the literature of possible universals across languages, which potentially underpin common underlying skills in learning to read. Verhoeven and Perfetti (2022) explored the universals and particulars in learning to read in seventeen languages representative of the five main writing systems internationally, including two of the four main Malaysian spoken languages (English and Chinese). They concluded that an overarching universal is that writing maps on to language, no matter the details of the orthography, creating a “common challenge in learning that mapping” (pg 161). They note that the particulars of the language and its orthography however impact the speed, efficiency, and manner of the learning. If we consider that mapping symbols to sound (orthography to phonology) or symbols to meaning (orthography to semantics) is a similar learning process, and one that is potentially universal across languages, then it is possible that Dynamic Reading Tasks (DRT) / SSP tasks would work in those reading morphosyllabic or syllabic orthographies, as well as those reading alphabetic orthographies. However, it is also possible that the particulars of these orthographies could mean that a different test of learning is

required, such as mapping symbols to meaning, or to different sizes of phonological information, such as syllables.

Therefore, our aim was to identify if Elbro et al.’s (2012) dynamic reading test could be a suitable diagnostic tool for reading difficulties in a simultaneous multilingual and multiliterate community in Malaysia; and as an auxiliary question, to explore the assertion that such a test would be more sensitive to those reading alphabetic languages as first literacy (Li1). We recruited a convenience sample of 59 Malaysian adults. Due to limitations in recruiting diagnosed dyslexic adults in Malaysia, we used a lexical decision task (LDT) in English, Chinese and Bahasa Melayu (BM) to identify reading proficiency across these three languages, measured phonological awareness in BM, and aimed to identify possible reading difficulties using the self-report Adult Reading History Questionnaire (ARHQ; Lefly & Pennington, 2000). Approximating Elbro et al.’s approach, we compared two reading proficiency groups based on English LDT reaction time (as accuracy was near ceiling) on the reading-related measures. We used English LDT as this appeared to be most participants’ first or second most proficient language based on self-report and LDT scores, and was also the language of current education for most participants. Given the lack of objective reading ability groups, we then performed a hierarchical linear regression to identify if reading skill (as measured by LDTs) or potential reading difficulty (ARHQ) were predicted by the dynamic reading test scores, after controlling for phoneme awareness. We then looked at whether those with an alphabetic First Literacy (Li1) showed different patterns of prediction to those with morphosyllabic Li1s, predicting that if Verhoeven and Perfetti (2022) were correct in their assertion that mapping writing to language is a universal skill in reading development, then there would be similar patterns across literacy groups. Language of first literacy was measured using the Language and Social Background Questionnaire (LSBQ; Anderson et al., 2018), modified to capture the multiple languages and literacies found in Malaysia. It was predicted both that the DRT would significantly predict LDT RTs and ARHQ scores taking phoneme awareness into account, and that the morphosyllabic Li1 group would show a similar pattern of results to those in the alphabetic Li1 group.

## Methods

### Participants

The study consisted of 59 participants, majority female, recruited via opportunity sampling ( $M_{age}=22.3$  years,  $SD=5.19$ ). Research credits were awarded to psychology students. Participants gave informed consent at the start of the study.

Participants' language backgrounds were complicated; the modified LSBQ described below identified first language (speech and literacy), self-reported proficiency and the Lexical Decision Tasks (described below) were used to determine the language in which participants were most accurate and quickest. Although the numbers appear relatively stable within-participants, there were noticeable differences between self-reported proficiency and that measured by the lexical decision tasks. Some participants had noticeable proficiency differences across the three languages, whilst others had more balanced patterns across measurement (see Table 1).

### Materials

#### Language and social background questionnaire (LSBQ; Anderson et al., 2018).

This questionnaire was modified to allow multiple languages and literacies to be reported. Nine questions were asked regarding up to six languages, about age of acquisition, proficiency, use across various

situations etc. This questionnaire typically results in a numeric value which represents the balance of bilingualism per participant. However, we used it to identify first / earliest spoken and written languages, and the spoken and written languages in which each participant felt most proficient.

The adult reading history questionnaire (ARHQ; Lefly & Pennington, 2000).

The English-language questionnaire measures attitudes towards reading, experience with literacy and numbers during childhood as well as family-based risk factors such as history of dyslexia or reading difficulty in the family. Participants were asked to rate each question or statement using a 5-point Likert scale (Range=0–4). Responses in between points were accepted. The score was calculated as the total ratings from the first 23 items divided by 92, giving a proportion score. Scores above 0.3 indicate a positive history of reading disability in US samples. The questionnaire was not amended for this sample and was used as a proxy for reading ability in light of the community sample and lack of potential participants with a formal diagnosis of reading difficulty.

Dynamic reading test (Elbro et al., 2012).

The experimenter provided instructions non-verbally, using hand signals and motions. First, participants were taught to associate 3 novel symbols to their sounds:  $\text{F}=/s/$ ,  $\text{D}=/m/$ ,  $\text{B}=/a/$  as in *calm*. Participants moved to the next phase after 3 consecutive

**Table 1** Distribution of participants according to language characteristics

	First spoken language	First literate language	Proficient spoken language	Proficient literate language	LDT accuracy	LDT RT
Bahasa Melayu	13	12	13	12	9	9
Chinese	33	27	32	30	22	14
English	13	20	14	17	22	36
Other						
BM+E					2	
C+E					3	
BM+C+E					1	

Reported first language and literacy and reported most proficient language and literacy from Language and Social Background Questionnaire (LSBQ), language of best performance in Lexical Decision Task (LDT) accuracy and reaction time

correct trials or 10 trials. In phase two, participants learned to read pairs of the previously learned symbols with corrective feedback. The experimenter first demonstrated by moving the single-letter cards from phase one together and producing the correct sounds for the sequence (e.g.  $\diamond + \square = [\text{ma}]$ ). Then, participants were invited to read 4 letter cards, each containing a pair of the previously learned symbols (e.g.  $\square \square = [\text{sa}]$ ), that were presented in a random order. There was a maximum of 5 trials. Participants moved onto phase three after reading all pairs correctly in two consecutive trials. Participants who did not meet the criteria did not progress to Phase 3 and automatically scored 0 as the DRT score. Phase three required participants to read twelve four-letter non-words comprised of the three symbols from phase one. The experimenter continued to provide corrective feedback but help with sound synthesis was not given if any errors were made by the participant. The score was the number of ‘words’ read correctly. Testing ended after 3 consecutive errors. Phase 3 scores were used in analyses, with higher scores representing better ‘word’ reading.

#### Phoneme awareness task

Phoneme awareness was measured with a phoneme counting task consisting of 24 common Bahasa Melayu words (see Appendix A). Each written word was presented on a card. Participants were asked to count the number of sounds and put a marker next to the card for each sound. For example, the word ‘bumi’ has the sounds /b/ /u/ /m/ /i/ (4 marks). The total of the correct responses was the score, with higher scores representing better performance.

#### Lexical decision tasks

Lexical Decision Tasks (LDT) in Chinese, Bahasa Melayu and English were prepared for this study. Characters from the Chinese Lexical Project for two-character compound words (Tse et al., 2017) were used for the Chinese lexical decision task. Word and non-word pairs were chosen from the database of 25,000+ word pairs by filtering for accuracy (0.97–1.00), log-transformed reaction time (bottom 10%), reaction time (< MeanRT=646.18) and frequency (top 5% of Cai and Brysbaert’s (2010) frequency measure for both raw subtitles and contextual

diversity). After trimming, 130 character pairs were chosen randomly (65 words and 65 non-words) and used for the lexical decision task.

Stimuli for the BM Lexical Decision task were obtained from the Malay Lexicon Project’s (Yap et al., 2010) database of over 9500 words. The sample was filtered for accuracy (0.97–1.00), reaction time (bottom 10%), and frequency (equal to or greater than the mean log frequencies,  $\geq 0.96$ ). 65 words were randomly chosen from the remaining pool. Non-words were not available from this database. Thus, 65 pseudowords were created using the chosen Malay words to match for bigram frequency. Each of the chosen Bahasa Melayu words was manually split into two parts, and the first part of one word was combined with the second half of another word. Each pseudoword was checked by two native Bahasa Melayu speakers and also cross-checked using the Malay Language dictionary, Kamus Dewan.

The English word stimuli were obtained from the English Lexicon Project (Balota et al., 2007) database, accessed online through [elexicon.wustl.edu](http://elexicon.wustl.edu). The sample was filtered for accuracy (0.97–1), log frequency (6.16), letter length (6–8) and reaction time (bottom 10%). 65 words were then randomly selected for our final sample and used to create another 65 pseudowords as per the Bahasa Melayu Lexical Decision Task. Similarly, all pseudowords were checked using the Cambridge Dictionary to ensure they were not real words.

Responses that were too fast or too slow (RT  $> 1.5$  s or  $< 0.2$  s) were excluded from the analyses. Reaction Time was calculated using data from accurate responses only. Higher Accuracy and RT scores represent higher proficiency.

The lexical decision tasks were administered to each participant in the same order: Chinese, followed by BM, then English. Participants were asked to decide as quickly as and accurately as possible if each word that appeared on the screen was a real word or not using the keyboard to respond. At the start of each lexical decision task, participants were given a practice trial consisting of 5 words and 5 non-words out of the chosen 130-word samples for each language. The words in the practice trials were not reused for the test trials; thus the test trials consisted of 120 words and non-words. Stimuli were presented as white characters in lowercase Arial font, letter height 0.1 in the center of the screen. A 0.5 s fixation

point ‘+’ at the centre of the screen was presented before each stimulus, followed by a 0.12 s interstimulus interval for all trials. Target words remained on screen until participants responded with either of the response keys.

### Procedure

A favorable opinion for conduct was granted by the University of Reading Malaysia research ethics committee in adherence with the tenets of the Declaration of Helsinki. Participants were briefed and given information sheets and consent forms at the start of the experiment. All participants were asked to complete the questionnaires and tasks in the same order. LSBQ was administered first, followed by ARHQ, DRT, Phoneme Awareness Task, Lexical Decision Tasks and a debrief at the end.

### Results

Phoneme Awareness, Dynamic Reading Test and Adult Reading History Questionnaire (ARHQ) scores significantly correlated with each other (see Table 2 below). Note that DRT scores correlate negatively as a high score indicates higher likelihood of reading difficulty, (i.e. potentially poorer reading). Lexical Decision (English) task scores did not significantly correlate with any measure, with LDT-Eng Accuracy scores close to ceiling (mean=0.96. SD=0.031).

Following a similar approach to Elbro et al. (2012), a *t*-test was conducted with median-split scores from the Lexical Decision Task (Reaction Time; see Table 3 below for Descriptive statistics). A significant difference was found for the Dynamic Reading Test ( $t(57)=2.25$ ,  $p=0.032$ ,  $d=0.57$ ), with those faster on the LDT-Eng scoring significantly

higher on the DRT than those slower on the LDT-Eng. No significant differences were found for any other reading measure (Phoneme Awareness:  $t(57)=1.05$ ,  $p=0.299$ ; Adult Reading History Questionnaire:  $t(57)=-1.42$ ,  $p=0.161$ ; Lexical Decision Task—English Accuracy:  $t(57)=1.19$ ,  $p=0.241$ ).

Two hierarchical linear regression analyses were conducted to predict Lexical Decision Task—English (RT; Model 1) and Adult Reading History Questionnaire scores (Model 2) with Phoneme Awareness entered as a predictor in Step 1 and Dynamic Reading Test added in Step 2 (see Table 4). When predicting LDT-English Reaction Times, Step 1 (Phoneme Awareness) was not significant ( $F(1,57)=0.137$ ,  $p=0.713$ ), and adding Dynamic Reading Task (DRT) scores (Step 2) was also nonsignificant ( $F(2,56)=1.603$ ,  $p=0.21$ ;  $R^2=0.04$ ). When predicting Adult Reading History Questionnaire scores, Phoneme Awareness (Step 1) was significant ( $F(1,57)=4.504$ ,  $p=0.038$ ),  $R^2=0.073$ , and when DRT scores were added (Step 2), this was also significant ( $F(2,56)=4.613$ ,  $p=0.014$ ),  $R^2=0.111$ , although only DRT remained a significant predictor ( $t=-2.109$ ,  $p=0.039$ ).

To further investigate potential differences between participants first reading alphabetic (English or BM) or morphosyllabic (Chinese) languages (Li1), we compared groups using a *t*-test (see Table 5). Participants who reported their Li1 as English or BM were in the Alphabetic group ( $n=32$ ), and those reporting their Li1 as a Chinese language / dialect were in the Morphosyllabic group ( $n=27$ ). No participants reported Tamil as Li1. No differences were found between Language of First Literacy groups for possibility of reading difficulty (ARHQ), Phoneme Awareness, or Dynamic Reading Test. However, there was a significant difference in accuracy on the Lexical Decision Task in English,

**Table 2** Correlations between reading-related measures

Variable	ARHQ	PA	LDT-Eng Accuracy	LDT-Eng RT
Dynamic Reading Test	– 0.33*	0.27*	0.24	– 0.23
ARHQ	–	– 0.27*	– 0.17	– 0.11
PA		–	0.18	– 0.05
LDT-Eng Accuracy			–	0.02

ARHQ Adult Reading History Questionnaire, PA Phonological Awareness, LDT Lexical Decision Task (English), RT Reaction Time

\*  $p < 0.05$ , two-tailed,  $N=59$

**Table 3** Reading-related measures for Slower and Faster readers

	Slower readers ( <i>n</i> =29)		Faster readers ( <i>n</i> =30)		<i>t</i> (57)	<i>p</i>	Cohen's <i>d</i>
	M	SD	M	SD			
PA	15.27	(7.81)	17.28	(6.85)	1.049	.299	.27
DRT	10.07	(4.07)	11.76	(0.64)	2.213	.031	.58
ARHQ	0.412	(0.113)	0.375	(0.083)	-1.415	.162	.37
LDT-Eng Accuracy	0.96	(0.037)	0.97	(0.022)	1.186	.241	.31

Reading speed grouped by median-split on Lexical Decision Task—English Reaction Times. *PA* Phoneme Awareness, *DRT* Dynamic Reading Test, *ARHQ* Adult Reading History Questionnaire, *LDT-Eng* Lexical Decision Task English

**Table 4** Hierarchical Regression Results for Reading (Lexical Decision Task and Adult Reading History Questionnaire)

		<i>B</i>	SE <i>B</i>	$\beta$	<i>R</i> <sup>2</sup>	$\Delta R^2$
Lexical Decision Task-English (RT)	Step 1				0.002	0.002
	Constant	0.695	0.026			
	Phoneme Awareness	-0.001	0.001	-0.049		
	Step 2				0.04	0.02
	Constant	0.752	0.042			
	Phoneme Awareness	0	0.001	0.016		
Adult Reading History Questionnaire	Step 1				0.07	0.07*
	Constant	0.454	0.031			
	Phoneme Awareness	-0.004	0.002	-0.271*		
	Step 2				0.11*	0.07
	Constant	0.535	0.049			
	Phoneme Awareness	-0.003	0.002	-0.196		
	Dynamic Reading Task	-0.009	0.004	-0.272*		

*N*=59. *RT* Reaction Time

\**p* < .05

with the morphosyllabic group scoring significantly lower ( $M=0.95$ ,  $SD=0.3$ ) than the Alphabetic group ( $M=0.97$ ,  $SD=0.03$ ;  $t(57)=2.56$ ,  $p=0.013$ ).

We then repeated the hierarchical linear regressions above with ARHQ (as this was significant for the whole group) for each Li1 group (see Table 6). When the linear regression was conducted on Li1-Alphabetic, Model 1 (PA only) was not significant ( $F(1,30)=2.484$ ,  $p=0.125$ ) but Model 2 (PA+DRT) was significant ( $F(2,29)=7.061$ ,  $p=0.003$ ). When the linear regression was conducted on Li1-Morphosyllabic, Model 1 ( $F(1,25)=1.71$ ,  $p=0.203$ ) and Model 2 ( $F(2,24)=0.824$ ,  $p=0.451$ ) were both non-significant.

## Discussion

Dynamic Reading Test scores correlated with Phoneme Awareness measures, a more typical assessment of skills underlying reading acquisition, and both significantly correlated with one measure of adult reading, the Adult Reading History Questionnaire (ARHQ; Lefly & Pennington, 2000), but not the more objective Lexical Decision Task in English, neither Accuracy nor Reaction Times. In addition, creating a rough reading proficiency measure using a median split of LDT-RTs, we found a significant group difference between slower and faster readers on the DRT measure. This somewhat aligns with Elbro et al.'s (2012) findings that dynamic test measures correlated significantly with reading-related

measures, although it should be noted that though significant, our correlations were small (~0.3) compared to those reported by Elbro et al. (~0.5 for DRT correlations; 2012). Hierarchical linear regression analyses across the whole sample demonstrated that when predicting ARHQ scores the addition of DRT significantly contributed to the model above PA, but neither significantly predicted LDT-RT scores. This indicates that the DRT could provide useful information regarding learning to read; however with its significant correlation with PA, the variance explained above PA on ARHQ could be an artifact, although there were no concerns about multicollinearity. The non-significant prediction of the LDT-RT scores, which is a more objective measure of proficiency (albeit in one particular language), also points to possible issues with the PA and DRT measures. There are therefore two possible interpretations of these data: both the ARHQ and DRT tap into an underlying aspect of learning to read, or the measures used are insensitive. Whilst we acknowledge the choice of measurements was far from ideal (see below), we believe this result indicates that some form of Dynamic Reading Test may be an appropriate measure for reading difficulties in multilingual populations, avoiding the difficulties of finding context-appropriate normed measures across various language profiles as described by Elbeheri and Everatt (2016).

No significant difference in dynamic task performance, phoneme awareness and ARHQ was found between the groups who acquired literacy in different orthographies, though there was a significant difference in LDT-English accuracy measures, with the Morphosyllabic group performing significantly poorer, albeit with scores close to ceiling. This might appear to support our assertion that language of first literacy (L1) may not impact the efficacy of the DRT, and demonstrate that this test could be suitable for readers across orthography. However, when groups were created by L1 orthography, there were different predictor patterns, with DRT significantly predicting ARHQ in the Alphabetic group, but not the Morphosyllabic group. It could also be interpreted that whilst learning mappings is a universal requirement (Verhoeven & Perfetti, 2022), the particulars of the languages are also necessary to consider when approaching testing for reading difficulties or dyslexia. It is clear that with the groups

containing only around thirty participants each, low power means we cannot draw any conclusions regarding the patterns in these data. In short, we have not adequately determined whether symbol-sound paradigms are differentially sensitive dependent upon orthographic experience.

There were several limitations regarding the measures used in this study. A key issue with Elbro et al.'s (2012) DRT approach is the effect of the stop rule, whereby participants who fail to reach criterion at Phase 2 receive a score of 0 overall. This leaves the data looking rather bimodal (though meeting assumptions for regression analysis), and could be an issue from a research perspective. Therefore, a different measure of symbol-sound matching, which gives a more normally distributed, continuous variable might be more appropriate to thoroughly explore the question of non-language-based reading assessment in multiliterate individuals. Further, our approach to phonological testing is problematic as it uses a written linguistic prompt in order to measure a phonological response (tapping phonemes). In the future, measures of phonological processing should be auditory only and tested in multiple or artificial languages as well as considering language-specific identifiers of dyslexia, such as fluency for transparent languages (e.g. Rapid Automatized Naming), visuospatial skills for logographic orthographies (e.g., visual attention span tests), etc.

If we assume that DRT-type tasks could be useful to detect reading difficulties in a multilingual context, the next steps require the use of a more accurate measure of reading ability than the ARHQ and lexical decision tasks. Whilst there was sensitivity in terms of variance of the ARHQ, it suffers from a similar issue of being delivered in one language, English, which may have culturally inappropriate items or be difficult for those with a reading disorder to understand in a potentially unsystematic manner across different language contexts. Indeed, we identified some items as perhaps not being culturally relevant in Malaysia, with items referencing reading books, magazines and newspapers for pleasure perhaps being understood differently in Malaysia compared to America, its country of origin. We therefore ran the analyses with and without these items, but found the differences were negligible and therefore used the full questionnaire. We also used the accuracy and reaction times of the English version of the Lexical

**Table 5** Age and Reading Measures Comparisons across Alphabetic and Morphosyllabic Groups

	Alphabetic (N=32)		Morphosyllabic (N=27)		t (57)	p	Cohen's <i>d</i>
	M	SD	M	SD			
Age	22	(3.1)	22.63	(6.67)	− 0.477	0.635	0.13
Dynamic Reading Test	11.03	(2.96)	10.74	(3.17)	0.364	0.717	0.1
Phoneme Awareness	16.88	(7.64)	15.52	(7.08)	0.702	0.486	0.18
ARHQ	0.39	(0.1)	0.40	(0.1)	− 0.558	0.579	0.15
LDT-Eng Accuracy	0.097	(.03)	0.95	(.03)	2.56	0.013	0.67
LDT-Eng RT	0.67	(.08)	0.7	(.08)	− 1.06	0.294	0.28

**Table 6** Linear Regression Results for Adult Reading History Questionnaire across Language Groups

Adult reading history questionnaire	B	SE B	B	R <sup>2</sup>	ΔR <sup>2</sup>
Alphabetic	Step 1			0.08	0.08
	Constant	0.448	0.042		
	Phoneme Awareness	− 0.004	0.002	− 0.277	
	Step 2			0.33**	0.25
	Constant	0.62	0.064		
	Phoneme Awareness	− 0.003	0.002	− 0.21	
Morphosyllabic	Step 1			0.06	0.06
	Constant	0.459	0.048		
	Phoneme Awareness	− 0.004	0.003	− 0.253	
	Step 2			0.06	0
	Constant	0.454	0.073		
	Phoneme Awareness	− 0.004	0.003	− 0.261	
	Dynamic Reading Task	0.001	0.007	0.017	

Alphabetic  $N=32$ ; Morphosyllabic  $N=27$ . Groups determined by self-reported language of first literacy

\*\*  $p < 0.01$

Decision task. Lexical decision tasks are not a gold standard measure of reading proficiency, given that they measure recall accuracy and speed rather than decoding as would be typical. In addition, choosing to use only one of the three performed could be taken as being Anglocentric ourselves. We analysed the data again, using the LDT of the language participants reported being most proficient in reading, and found the same patterns as in English, which for all our participants is the language used in their work / study context. We find ourselves in an unenviable situation: how do we judge the utility of a potential measure of reading difficulty, in a scenario where it is exceedingly difficult to identify with accuracy those with reading difficulty. One approach would be to recruit adults already diagnosed with dyslexia, using only

those who have been tested in more than one language. We suspect there are few such cases available, though the increase of interest in this area and the increased availability of better local language tests (e.g. Lee et al., 2020) might yield more stable diagnoses in the future. Another method would be to take a developmental approach, testing pre-readers or those in the early years of schools, to identify the predictive power of this approach, following a similar approach to Horbach et al.'s (2015, 2018) series of studies.

Finally, we categorised language groups on the basis of self-reported L1 through an amended LSBQ (Anderson et al., 2018), following the idea proposed in the Psycholinguistic Grain Size Theory (Ziegler & Goswami, 2005) that the language of first literacy

affects subsequent literacy strategies. However, given the simultaneous rather than sequential nature of multilingualism in Malaysia, this approach may not be appropriate. We asked participants for information on their first spoken language, and also what they now consider their most proficient language and literacy to be. Only three participants reported they were most proficient in a literacy that was not their first literacy (all with L1 English but most proficient now in Mandarin), but our somewhat simplistic approach to language context may have led to a decreased sensitivity in our measures. Better conceptualisation of these groups may lead to different conclusions. Interesting future approaches may be Gullifer and Titone's (2020) Language Entropy measure, which captures the balance of language use across multiple contexts, as does Li et al.'s (2006) Language History Questionnaire approach. An alternate approach would be to use a DRT measure in morphosyllabic monolingual children, such as Japanese or Chinese readers, to investigate the predictive power of this measure. To our knowledge, this has not yet been done in such groups.

In summary, we have demonstrated that dynamic testing in an artificial orthography could be used as an appropriate measure of reading difficulty in adult multilingual readers, including simultaneous rather than sequential multilingualism, and with complex language contexts. To further investigate this, young children in a simultaneous language context should be followed longitudinally, though concerns remain regarding objective measures of reading ability in this sample. Our results also imply that, in agreement with Elbro et al.'s (2012) statement, there could be subtle differences in the sensitivity of this measure with readers from non-alphabetic literacy backgrounds, though the complex language contexts of our participants mean we cannot draw any firm conclusions. Future research should use this measure in morphosyllabic monoliterate/linguals, which will give a stronger test of Elbro et al.'s (2012) assertion than our current sample can provide.

## Appendix A: Phoneme awareness task words

rindu	soal	minat
jahat	teman	sempit
cepat	masuk	kawan
niat	kerja	kaya
pasti	ingat	pondok
sahaja	bumi	rasa
tanah	tentu	sampai
dapat	khlas	bayar

**Author contributions** RP and CHH designed the study. CHH prepared stimuli and collected data. RP and CHH analysed the data. RP wrote the main manuscript text. All authors reviewed the manuscript.

**Funding** Not applicable.

**Data availability** Available on request.

**Code availability** Not applicable.

## Declarations

**Competing interests** The authors declare no competing interests.

**Conflicts of interest** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Aaron, P. G., Joshi, R.M., Boulware-Gooden, R., & Bentum, K. (2008). Diagnosis and treatment of reading disabilities based on the Component Model of Reading: An alternative to the discrepancy model of learning disabilities. *Journal of Learning Disabilities*, 41, 67–84. <https://doi.org/10.1177/0022219407310838>

Anderson, J. A., Mak, L., Chahi, A. K., & Bialystok, E. (2018). The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, 50(1), 250–263. <https://doi.org/10.3758/s13428-017-0867-9>

Aravena, S., Snellings, P., Tijms, J., & van der Molen, M. W. (2013). A lab-controlled simulation of a letter–speech sound binding deficit in dyslexia. *Journal of Experimental Child Psychology*, 115(4), 691–707. <https://doi.org/10.1016/J.JECP.2013.03.009>

Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2016). Predicting responsiveness to intervention in dyslexia using dynamic assessment. *Learning and Individual Differences*, 49, 209–215. <https://doi.org/10.1016/j.lindif.2016.06.024>

Aravena, S., Tijms, J., Snellings, P., & van der Molen, M. W. (2018). Predicting individual differences in reading and spelling skill with artificial script-based letter–speech sound training. *Journal of Learning Disabilities*, 51(6), 552–564. <https://doi.org/10.1177/0022219417715407>

Cai, Q. & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, 5(6), p.e10729. <https://doi.org/10.1371/journal.pone.0010729>

Cambridge University Press (n.d.). Cambridge Dictionary. Accessed October 22nd 2018 from <https://dictionary.cambridge.org/>

Chikamatsu, N. (1996). The effects of L1 orthography on L2 word recognition: A study of American and Chinese learners of Japanese. *Studies in Second Language Acquisition*, 18(4), 403–432. <https://doi.org/10.1017/S0272263100015369>

Daniels, P.T. & Share, D.L. (2018). Writing system variation and its consequences for reading and dyslexia. *Scientific Studies of Reading*, 22(1), 101–116. 10.1080/10888438.2017.1379082

Department of Statistics Malaysia. (2020). *Current Population Estimates, Malaysia*. Retrieved July 25 2020, from [https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul\\_id=OVByWjg5YkQ3MWFZRTN5bDJiaEvhZz09&menu\\_id=L0pheU43NWJwRWVSZkIWdzQ4TlhUUT09](https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=155&bul_id=OVByWjg5YkQ3MWFZRTN5bDJiaEvhZz09&menu_id=L0pheU43NWJwRWVSZkIWdzQ4TlhUUT09)

Dewan Bahasa dan Pustaka (n.d.). *Kamus Dewan*. Retrieved October 22<sup>nd</sup> 2018 from <http://pprm.dbp.gov.my/>

Dzalani, H., & Shamsuddin, K. (2014). A Review of Definitions and Identifications of Specific Learning Disabilities in Malaysia and Challenges in Provision of Services. *Pertanika Journal of Social Sciences & Humanities*, 22(1). <http://www.pertanika.upm.edu.my/pjssh/browse/regular-issue?article=JSSH-0620-2012>

Elbeheri, G., & Everatt, J. (2016). Principles and guidelines in test construction for multilingual children. In L. Peer & G. Reid (Eds.), *Multilingualism, Literacy and Dyslexia* (pp. 49–60). Routledge.

Elbro, C., Daugaard, H. T., & Gellert, A. S. (2012). Dyslexia in a second language?—a dynamic test of reading acquisition may provide a fair answer. *Annals of Dyslexia*, 62(3), 172–185. <https://doi.org/10.1007/s11881-012-0071-7>

Gomez, C. (2004). Dyslexia in Malaysia. By I. Smythe, J. Everatt & R. Salter (Eds.) *International book of dyslexia: A guide to practice and resources* (pp158–163). John Wiley & Sons.

Gullifer, J.W. & Titone, D. (2020). Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, 23, 283–294. <https://doi.org/10.1017/S1366728919000026>

Horbach, J., Scharke, W., Cröll, J., Heim, S., & Günther, T. (2015). Kindergarteners' performance in a sound-symbol paradigm predicts early reading. *Journal of Experimental Child Psychology*, 139, 256–264. <https://doi.org/10.1016/j.jecp.2015.06.007>

Horbach, J., Weber, K., Opolony, F., Scharke, W., Radach, R., Heim, S., & Günther, T. (2018). Performance in sound-symbol learning predicts reading performance 3 years later. *Frontiers in Psychology*, 9, 1716. <https://doi.org/10.3389/fpsyg.2018.01716>

Jalil, S. B., & Liow, S. J. R. (2008). How does home language influence early spellings? Phonologically plausible errors of diglossic Malay children. *Applied Psycholinguistics*, 29 (4), 535–552. <https://doi.org/10.1017/S0142716408080235>

Kormos, J. (2017). The effects of specific learning difficulties on processes of multilingual language development. *Annual Review of Applied Linguistics*, 37, 30–44. <https://doi.org/10.1017/S026719051700006X>

Lachmann, T., Bergstrom, K., Huber, J. & Nuerk, H.C. (2022). Diagnosis of dyslexia and dyscalculia: Challenges and controversies. In M.A. Skeide (Ed.). *The Cambridge Handbook of Dyslexia and Dyscalculia* (pg. 383–409). Cambridge University Press.

Lee, L. W. (2008). Development and validation of a reading-related assessment battery in Malay for the purpose of dyslexia assessment. *Annals of Dyslexia*, 58(1), 37–57. <https://doi.org/10.1007/s11881-007-0011-0>

Lee, J. A., Lee, S., Yusoff, N. F. M., Ong, P. H., Nordin, Z. S., & Winskel, H. (2020). An early reading assessment battery for multilingual learners in Malaysia. *Frontiers in Psychology*, 11, 1700. <https://doi.org/10.3389/fpsyg.2020.01700>

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, 33(3), 286–296. <https://doi.org/10.1177/002221940003300306>

Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behaviour Research Methods*, 38(2), 202–210. <https://doi.org/10.3758/bf03192770>

Lim, I. (2017, July 3). What you should know about Chinese schools in Malaysia. *Malay Mail*. <https://www.malaymail.com/news/malaysia/2017/07/03/what-you-should-know-about-chinese-schools-in-malaysia/1412233>

Lindgrén, S. A., & Laine, M. (2007). The adaptation of an adult group screening test for dyslexia into Finland-Swedish:

Normative data for university students and the effects of language background on test performance. *Scandinavian Journal of Psychology*, 48(5), 419–432. <https://doi.org/10.1111/j.1467-9450.2007.00592.x>

McBride-Chang, C., Liu, P. D., Wong, T., Wong, A., & Shu, H. (2012). Specific reading difficulties in Chinese, English, or both: Longitudinal markers of phonological awareness, morphological awareness, and RAN in Hong Kong Chinese children. *Journal of Learning Disabilities*, 45(6), 503–514. <https://doi.org/10.1177/0022219411400748>

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin*, 140(2), 409. <https://doi.org/10.1037/a0033890>

Ministry of Education Malaysia (2013). *Malaysian Education Blueprint 2013–2025 (Preschool to Post-Secondary Education)*. Putrajaya: Kementerian Pendidikan Malaysia.

Tse, C. S., Yap, M. J., Chan, Y. L., Sze, W. P., Shaoul, C., & Lin, D. (2017). The Chinese Lexicon Project: A megastudy of lexical decision performance for 25,000+ traditional Chinese two-character compound words. *Behavior Research*, 49, 1503–1519. <https://doi.org/10.3758/s13428-016-0810-5>

Verhoeven, L., & Perfetti, C. (2022). Universals in learning to read across languages and writing systems. *Scientific Studies of Reading*, 26(2), 150–164. <https://doi.org/10.1080/10888438.2021.1938575>

Wang, M., Koda, K., & Perfetti, C. A. (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition*, 87(2), 129–149. [https://doi.org/10.1016/s0010-0277\(02\)00232-9](https://doi.org/10.1016/s0010-0277(02)00232-9)

Yap, M., Liow, S. J. R., Jalil, S. B., & Faizal, S. S. B. (2010). The Malay lexicon project: A database of lexical statistics for 9,592 words. *Behavior Research Methods*, 42, 992–1003. <https://doi.org/10.3758/BRM.42.4.992>

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, 131(1), 3. <https://psycnet.apa.org/doi/https://doi.org/10.1037/0033-2909.131.1.3>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.