

An Exploration of Assessment of Early English Language Development for Kindergarten Children in Hong Kong

CHAN, Lydia Lai Seng

Department of Educational Studies, University of Oxford

SYLVA, Kathy

Department of Educational Studies, University of Oxford

Abstract: Research into Early Years assessment for preschoolers is a relatively new field of endeavour in Hong Kong, and there is a growing need for the identification and development of appropriate English language assessments for use with such children. This study explores three widely used Early English language measures developed for use with preschool children in U.K. (British Picture Vocabulary Scale II, Bryant & Bradley Phonological Awareness Assessment and Marie Clay Letter Identification Test), and investigates whether they are appropriate for the assessment of ESL children in Hong Kong, by evaluating their validity and sensitivity. The selected L1 measures were administered to 75 normally developing 4-year-old children (mean age = 4;6; SD = 5.89) from a bilingual Kindergarten in Hong Kong. The findings suggest that all three English language measures discriminated between the children with acceptable levels of sensitivity by yielding a range of normally distributed scores. Adequate evidence of concurrent criterion-related validity was also obtained through correlation analyses of the children's English test scores and their performance on a nonverbal cognitive assessment (Pattern Construction subscale of the British Ability Scales II) and relevant teacher ratings of their English. Furthermore, the children's vocabulary performance (BPVS-II) was found to be comparable to the U.K. EAL norming sample, and the possible effect of cultural bias was not substantial enough to decrease the overall validity of the instrument for this population. In short, it was found that, contrary to the research literature, the selected L1 English language measures are appropriate for assessing L2 English language skills of Hong Kong preschoolers, at least for those in the sample. This may suggest that the current popular case against the use of standardized tests in L2 assessment is not as unequivocal as may be commonly assumed.

Keywords: assessment, L2, preschool, English, vocabulary

Objectives

Although there has been considerable research on second-language acquisition among young children, there is very little systematic research on the appropriate assessment of their second-language abilities, especially in terms of identifying and developing developmentally and culturally appropriate assessment instruments. It is important to determine what level, if any, of proficiency these young children have in English, to diagnose their strengths and areas for improvement, and to keep track of their progress in acquiring the language. Appropriate language assessment, whether informal, classroom based, or large-scale, thus has a critical role to play in gathering the information for such purposes. Furthermore, the current absence of

such assessment tools impedes attempts to conduct programme evaluations and experimental interventions, as there are no valid means of rigorously measuring their impact.

The motivation for this study was a direct response to the lack of objective and validated measures of Early English Language development for assessing ESL preschool children in Hong Kong. This study, therefore, explores three widely used L1 English language measures developed for use with preschool children in the U.K., and investigates whether they could be appropriate for the assessment of ESL children in Hong Kong. The objective was not to compare Hong Kong children with their British counterparts, but rather the focus was on the validity and sensitivity of the instruments for Second Language (L2) learners within the Hong Kong context, which were evaluated both qualitatively and quantitatively. It was recognized that there may be various pitfalls of assessing L2 learners on tests developed for native speakers (e.g. content bias), but it is important to first field-test these instruments in order to decide whether or not to adapt these tools or to develop an alternative scale altogether.

Theoretical Framework

Assessing the English language development of culturally and linguistically diverse young children in a non-discriminatory manner is particularly difficult, due to the lack of appropriate assessment tools for both established and less well-established minority populations (Washington & Craig, 1999). In general, two major strategies have emerged to address this problem: (i) firstly, to develop alternative nonstandardized assessment measures, and (ii) secondly, to modify widely used standardized language instruments in an effort to reduce bias. Whereas the ongoing successful development of nonstandardized measures has been encouraging, the absence of standardized instruments for use with these children is a source of concern for those who adopt a 'balanced' multi-method, multi-source approach to assessment (Bracken, 1994; Dockrell, 2001; Nagle, 2000; Salinger, 2003; Salvia, Ysseldyke, & Bolt, 2006; Washington & Craig, 1999; Yaden, Rueda, Tsai, & Esquinca, 2004).

Much has been written about the pitfalls of assessing second language learners on standardized tests developed for the majority group, first language learners (Barona & Santos de Barona, 2000; P. H. Johnston & Rogers, 2003). It is believed that right from the design phase in the test development process, learner characteristics and the expected performance (the construct to be assessed) are more likely to be invalid for second language learners, thus creating content bias in tests (Garber & Slater, 1983; McLaughlin, Gesi Blanchard, & Osanai, 1995). This belief has been so widespread and deeply entrenched that there has virtually been no research study on L2 or bilingual learners using L1 standardized assessment instruments, especially for very young children.

Since most of the research in this field had been conducted on L2/EAL minority populations in the U.S. and U.K., the findings may or may not be directly transferable to the ESL preschool population in Hong Kong, and the only way to find out is through empirical

studies. Therefore, this small-scale exploratory study was undertaken to ‘test the waters’, without in any way dismissing the validity of past research.

Methodology

Three well-established and frequently used L1 English language measures developed in U.K. were carefully selected for field-testing on a convenience sample of 4-year-old children from a bilingual Kindergarten in Hong Kong. These individually administered measures test different aspects of children’s emergent English literacy skills (vocabulary, phonological awareness, and letter-name knowledge), and were judged to be psychometrically sound for their intended population of British children. They are all child-friendly and cater for a wide range of abilities, allowing for both verbal and nonverbal responses. Furthermore, they can be quickly and easily administered and rapidly scored with minimal formal training.

Before the selection of assessment measures was finalized, they were pilot tested on 10 children from the Kindergarten, who all seemed to enjoy the ‘games’ and had no difficulty interacting in English.

(i) British Picture Vocabulary Scale-II

Several aspects of children’s language skills are important at different points in the process of literacy acquisition, and initially, vocabulary is important (Whitehurst & Lonigan, 1998). One of the most well-established and generally accepted vocabulary tests in the U.S. is the Peabody Picture Vocabulary Test (PPVT). The British Picture Vocabulary Scale (BPVS-II) is strongly linked with the PPVT, and is now also widely recognized as a valuable assessment instrument for educational, clinical and research purposes in Britain. The BPVS-II is an individually administered, norm-referenced, wide-range test of hearing vocabulary for Standard English, and clear evidence is provided for its reliability and validity (Dunn, Dunn, Whetton, & Burley, 1997). The test contains four training plates, followed by 14 sets of 12 test items, which are arranged so that each successive set is more difficult than the preceding one. Each item has four simple black and white illustrations on a plate arranged in a two-by-two array. The subject’s task is to select the picture considered to illustrate best the meaning of a stimulus word presented orally by the examiner; hence it is a multiple-choice task.

Although the BPVS-II is normed for the British population, new local norms on pupils for whom English is an additional language (EAL) are now provided in a Technical Supplement (from the ages of 3;0 to 8;5). As noted in the Supplement, for EAL pupils the scale should only be viewed as a measure of level of attainment in English hearing vocabulary, and not as a measure of scholastic aptitude. Note that although some researchers have used the PPVT to assess bilingual and ESL children (Bialystok, Luk, & Kwan, 2005; Chow & McBride-Chang, 2003), the BPVS-II (with EAL norms) was chosen for this study instead because the British

test content might be a better match for the Hong Kong population, which continues to be influenced by its colonial history.

(ii) Bryant & Bradley Phonological Awareness Assessment

Phonological awareness refers to one's ability to represent spoken language as comprising discrete and recurrent sound elements (including phonemes, syllables, and words) (Justice, Invernizzi, & Meier, 2002). It is one of the most powerful predictors of later reading achievement (Bradley & Bryant, 1983; Bryant, Maclean, Bradley, & Crossland, 1990; Catts, Fey, Zhang, & Tomblin, 2001). Developing gradually during the preschool and early elementary period, children progress along a continuum representing shallow to deeper levels of awareness (Stanovich, 2000). Early attainments in phonological awareness include comprehending and producing rhyme and alliteration at the whole-word level and recognizing the intra-syllabic boundaries of words (Lonigan, Burgess, Anthony, & Barker, 1998).

Although it has never been formally published, the Bryant & Bradley Phonological Awareness Assessment has been widely used for research purposes by the *Effective Provision of Preschool Education* (EPPE) in U.K., and also in the evaluation of the Peers Early Education Partnership (PEEP) (Evangelou, Brooks, Smith, & Jennings, 2005). It is a quick test to administer, and Bryant and Bradley (1985) have demonstrated that children's scores on the initial rhyming tests are a strong predictor of their later progress (Bryant & Bradley, 1985). The test consists of two different sub-scales: Rhyme and Alliteration. The Rhyme subscale is presented as a game about "words that sound the same", and several examples of rhyming words are given to illustrate the notion of 'rhyme' at the beginning (e.g. hump and lump). Then 10 sets of 3 picture cards are presented one at a time, and the child is required to identify the words that sound the same or pick the odd one out (e.g. sail, nail, boot). As with the Rhyme subscale, Alliteration is presented as a game about "words that sound the same at the beginning". Again, there are 10 sets of 3 picture cards, and the child is required to identify the words that have the same beginning sound or pick the odd one out (e.g. cat, car, hen).

(iii) Marie Clay Letter Identification Test

Children's knowledge of individual letter names has also been identified as one of the foremost predictors of later reading achievement (Blatchford, Burke, Farquhar, Plewis, & Tizard, 1987; Catts, Fey, Zhang, & Tomblin, 2001; R. S. Johnston, Anderson, & Holligan, 1996). There is some evidence to suggest that explicit awareness of phonemes develops subsequent to children developing accurate representations and names of individual alphabet letters (R. S. Johnston, Anderson, & Holligan, 1996), thereby asserting the primacy of the alphabetic principle in children's early literacy development.

The Marie Clay Letter Identification test is designed to assess which letters the child knows (Clay, 1972). All letters, both lower case and capital, are presented to the child in random order, which should take 5 to 10 minutes. The child can respond by (i) naming the letter, (ii)

sounding the letter, or (iii) producing a word that begins with the same letter (e.g. ‘a’ for Apple).

Method

Validity refers to the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses of tests (AERA, 1999), and is therefore the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations (Salvia, Ysseldyke, & Bolt, 2006). There are various methods of validating test inferences, and there are many different types of evidence to examine (e.g. reliability, method of measurement, adequacy of norms etc.). Given the constraints of time and resources, a systematic empirical validation exercise was clearly unfeasible, so a small scale study was conducted to gather evidence on content validity and concurrent criterion-related validity, which were examined through a correlational study.

The nature of the criterion measure is extremely important, as the criterion itself must be valid if it is to be used to establish the validity of another measure (Salvia, Ysseldyke, & Bolt, 2006). Unfortunately quite often, as in the present case, no test which is known to be valid and reliable is available for the purposes of concurrent validation. Yet one does wish to know how the experimental tests compare with other measures that are known and used in that particular context, even though their reliability and validity are unknown. The less-than-perfect criterion measures selected for this study were the Pattern Construction subscale of the British Ability Scales-II (BAS-II), as a non-verbal cognitive measure, and relevant teacher ratings of the children’s English language ability. It is recognized that the results of any correlation must be treated very cautiously indeed, and a high correlation might not be expected partly because of the possible unreliability and uncertain validity of the criterion measures (Alderson, Clapham, & Wall, 1995).

Main Research Questions:

1. How appropriate, in terms of test content, are selected L1 Early English language measures for assessing L2 English language skills in preschool children in Hong Kong? [i.e. The themes, wording, and format of the items, tasks, or questions on a test, as well as the guidelines for procedures regarding administration and scoring (AERA, 1999)]
2. Do the selected language measures discriminate between Hong Kong ESL children with acceptable levels of sensitivity (i.e. tests should yield a range of performances from well above average to below average that closely approximates a standard normal distribution)?
3. Is there a significant relationship between the children’s scores on the selected language measures and their performance on a nonverbal cognitive measure?
4. Is there a significant relationship between the children’s scores on the selected language measures and teacher ratings of their English language ability?

Further Questions for BPVS-II:

- (i) To what extent are the published U.K. (EAL) norms similar to a local sample of ESL children in Hong Kong?
- (ii) To what extent does the instrument have content validity, based on item analysis data from a local sample of ESL children in Hong Kong?

Sample

The defining criteria set for participation were simply that the children had to be typically developing, chronologically aged 3;9 to 5;3, and had enrolled at the bilingual Kindergarten for at least 3 months. The resulting convenience sample consisted of 75 normally developing 4-year-old boys ($n = 41$; 55%) and girls ($n = 34$; 45%), with a mean age of 4;6 ($SD = 5.89$) from 12 different classes in the Kindergarten. About half of them were still in the 3-year-old Year Group ($n = 37$; 49%), while the other half were in the 4-year-old Year Group ($n = 38$; 51%). Their parents gave written voluntary informed consent on behalf of the children, and they also completed a questionnaire designed to collect demographic data.

Results

All three English language measures were deemed to be appropriate through qualitative analysis, in terms of test content. None of the children had any apparent difficulties with the themes, wording, and format of the items or tasks on the tests, and the administration guidelines were child-friendly. The BPVS-II (EAL) (mean = 107.00; $SD = 14.51$) and the Phonological Awareness Assessment (mean = 11.08; $SD = 7.04$) both discriminated between the children with acceptable levels of sensitivity by yielding a range of scores that were normally distributed (See Figures 1 & 2). The Letter Identification test of all 52 upper and lower case letters was found to be less sensitive, as the children's raw scores were non-normal and negatively skewed (mean = 40.31; $SD = 26.63$), but the distribution was successfully corrected by log transformation (See Figure 3). Since most of the children tested knew their alphabet quite well already, this particular test may be more discriminating for a younger age group.

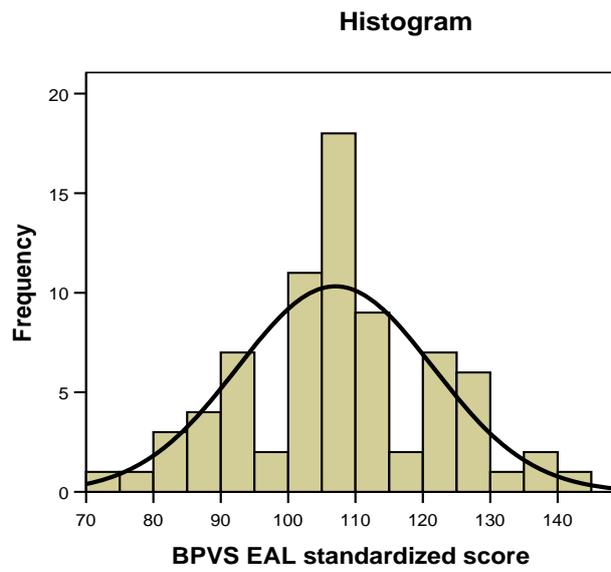


Figure 1: *Distribution of BPVS-II EAL Standardized Scores*

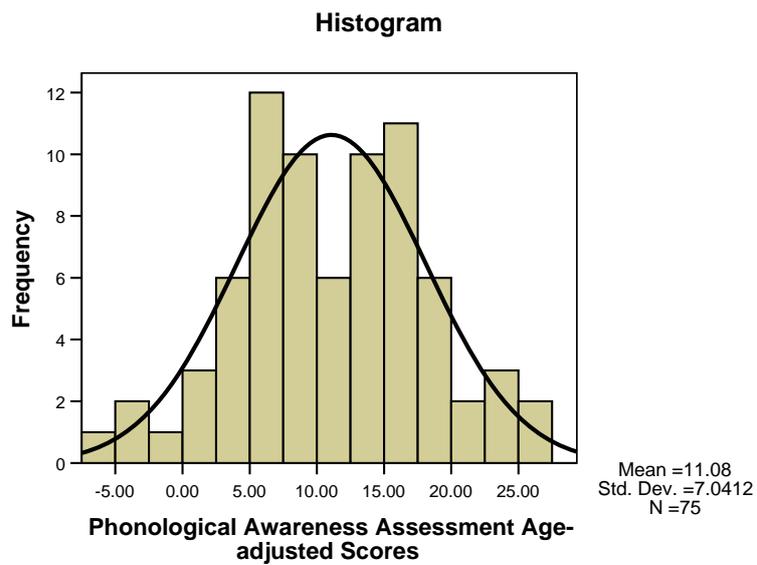


Figure 2: *Distribution of Phonological Awareness Age-adjusted Scores*

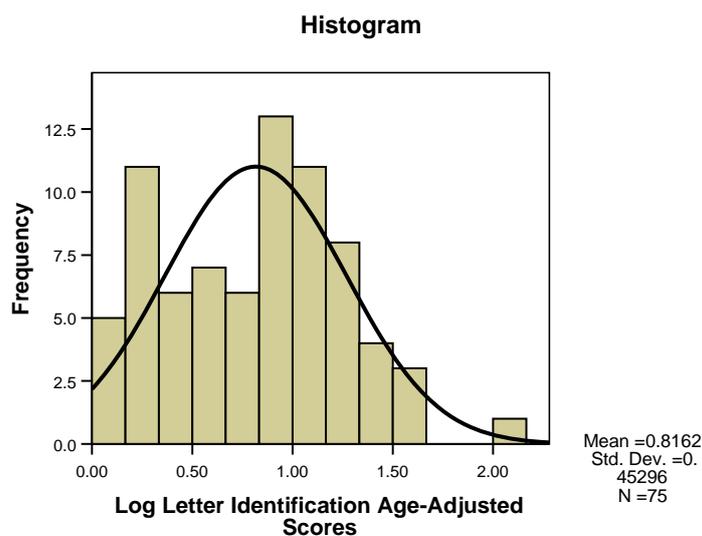


Figure 3: *Distribution of Log Letter Identification Age-adjusted Scores*

The children's non-verbal cognitive performance, as measured by the BAS Pattern Construction subscale (mean = 101.19; $SD = 18.36$), was significantly correlated with their scores on the Letter Identification test ($r = .34, p < .01$) and the Phonological Awareness Assessment ($r = .26, p < .05$). It was, however, non-significantly related to their BPVS-II EAL scores, which was mainly attributed to the children's variable exposure to English in the home, and potential factors uniquely inherent to second language acquisition that are nonexistent in monolinguals. Note that in a similar study previously conducted, the performance of 59 at-risk, African American preschoolers on PPVT-III was examined, and it found that there was no correlation between their PPVT-III scores and performance on a nonverbal cognitive measure. The authors explained that nonverbal cognitive tests are theoretically designed to examine cognitive ability without the influence of language, and the lack of significant relationship between scaled scores on the Triangles subtest of the Kaufman Assessment Battery for Children (KABC) and the PPVT-III suggested that this subtest does in fact assess this discrete functioning (Washington & Craig, 1999).

The other criterion measure used in this study was the English language teacher ratings on a 4-point scale (mean = 3.27; $SD = .56$), which unfortunately were rather insensitive (i.e. skewed towards the top end), especially for the 3-year-old Year Group. Unsurprisingly, therefore, while all three English language measures were found to be significantly correlated with the teacher ratings for the 4-year-old Year Group [BPVS-II: $\tau = .35, p < .01$; Phonological Awareness: $\tau = .29, p < .05$; Letter ID: $\tau = .41, p < .01$], only Letter Identification scores were significantly correlated for the 3-year-old Year Group ($\tau = .28, p < .05$), and diminished coefficients were found at the whole sample level [BPVS-II: $\tau = .19, p < .05$; Letter ID: $\tau = .20, p < .05$]. Nevertheless, it was felt that adequate evidence of concurrent criterion-related validity was obtained.

Perhaps the most interesting and encouraging findings were revealed in the further empirical analyses of the BPVS-II. It showed that the performance of the sample of ESL children in Hong Kong (mean = 107; *SD* = 14.5) were actually very similar, and even slightly superior, to the U.K. EAL norming sample (mean = 100; *SD* = 15). Although the classical item analyses of item difficulty and discriminability indicated that a few items in the scale might be culturally biased, the effect was clearly not substantial enough to decrease the overall validity of the instrument for this population (See Figure 4 & 5).

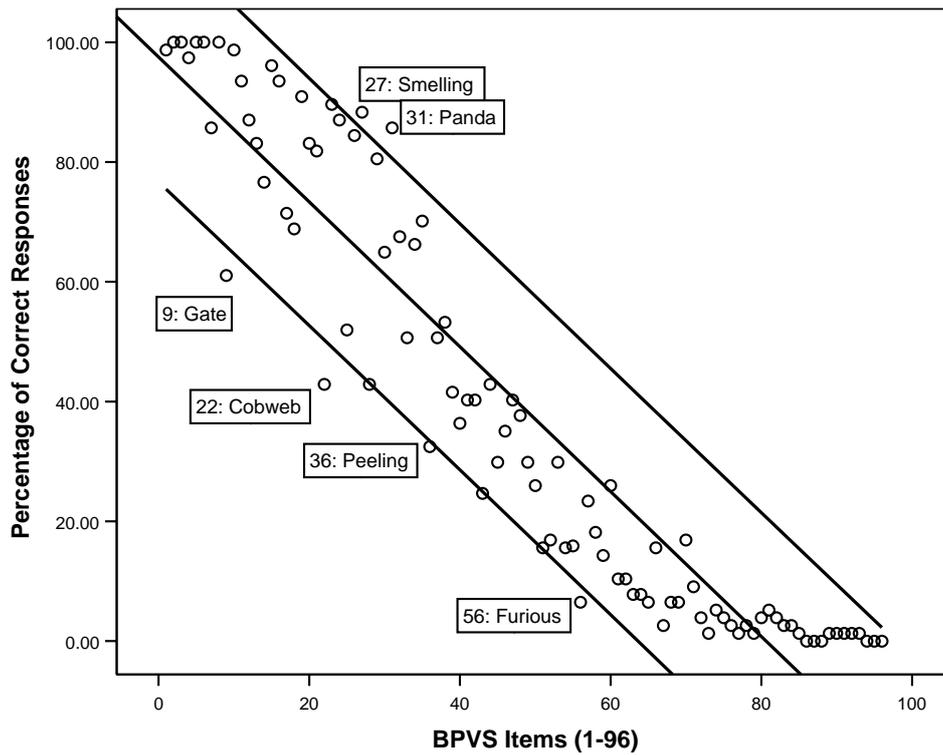


Figure 4: *Progression of item difficulty of the BPVS-II with 95% confidence interval*

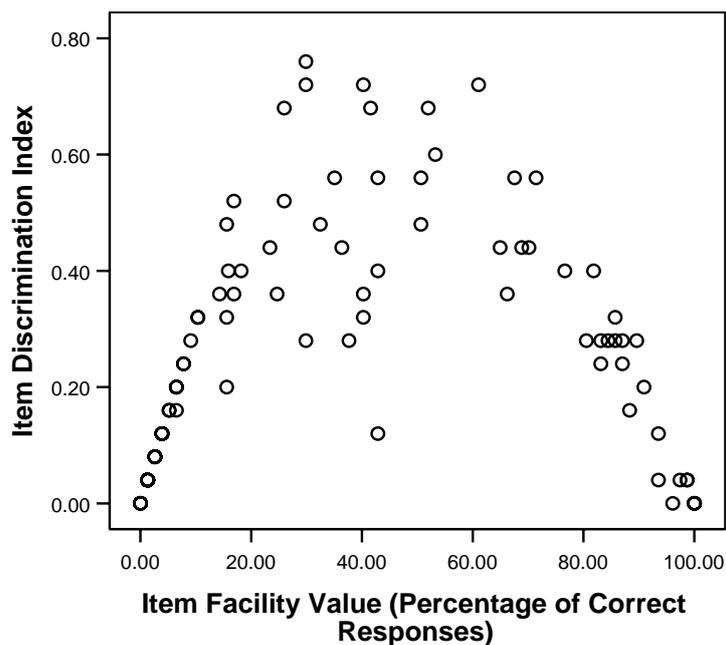


Figure 5: Relationship between item difficulty and item discriminability of the BPVS-II

In short, the general finding of the study, contrary to the research literature, is that the selected L1 standardized assessment measures are appropriate for assessing the L2 English language development of Hong Kong preschoolers, at least for those in the sample. This may suggest that the current popular case against the use of standardized tests in L2 assessment is not as unequivocal as is commonly assumed. Note, however, that their use should be recommended *only* as part of a profile of assessments, which includes qualitative measures and input from parents.

Suggestions for Future Research

There is much room for future research in both the validation and development of appropriate assessment of L2 English language development for Hong Kong children. Just building upon this small-scale study alone, the sample size could be significantly boosted to enable more rigorous and sophisticated statistical analyses to be done. If generalization of the research findings to the wider population is judged to be an important objective, then a representative sample must be obtained. More importantly, the appropriateness of the selected criterion measures should be thoroughly assessed beforehand to evaluate their reliability and validity. For the English language teacher ratings in particular, a new uniform rating scale with a corresponding scoring rubric might need to be constructed in close consultation with the Kindergarten principals and teachers. If teachers and assessors are to mark in a consistent manner, it is essential that they agree on the meaning and application of the criteria, so effective training and moderation sessions will be crucial.

Also, apart from gathering evidence related to concurrent criterion-related validity of the tests, an evaluation of their predictive criterion-related validity is equally important. Predictive validity refers to how accurately a person's current performance estimates the person's performance on a criterion measure administered at a later time, e.g. Does knowledge of a student's scores on these preschool English language tests allow an accurate estimation of the student's later English achievement in early Primary school? In addition, various measures of reliability should be obtained in order to provide a more comprehensive picture of the tests' psychometric strength.

Furthermore, while there has been much research conducted on school and home influences on monolingual children's emergent literacy development, there seems to be comparatively few studies on bilingual or ESL children. The findings of this present study suggest the importance of English exposure in both preschool settings and in the home, especially for children's L2 vocabulary development. Future studies might therefore develop systematic ways to measure such exposure, and more detailed data on home literacy practices as well as preschool experiences are required to provide additional insight. The children's maintenance of their first language (i.e. Chinese), for example, may be influential in their lexical development in English. The results of such investigations may have implications for instructional purposes and for guidance to parents of young bilingual children. Clearly, there is much room and need for further in-depth research.

Implications for Practice

Although the undertaking of this study was not directly intended to influence the practice of assessing L2 English language development of preschoolers in Hong Kong, the research findings do suggest that some L1 Early English language measures could be appropriate for this population. All three selected measures were found to be child-friendly (and teacher-friendly) tools that can generate much valuable information effectively and efficiently. They were conducted in a manner that was, although standardized, more like the normal literacy activities in which children engage in. The BPVS-II, in particular, could potentially be adapted and re-standardized with relatively minor alterations of certain test items.

It is important to note that these instruments are all oral tests of English language, which represent a significant departure from the traditional paper-and-pencil examinations that are used in some local Kindergartens. The Education and Manpower Bureau quite rightly criticises the use of such written tests, as they are developmentally inappropriate and imbalanced (Tong & Clem, 2006). Although the use of oral tests are likewise discouraged by the Bureau (CDC, 2006), the ones presented in this study do not seem to impose undue pressure on the children (who viewed them as fun games instead), and do not require them to memorise or regurgitate knowledge. Whether or not these tests should be widely used in Kindergartens, however, depends crucially on the goals and objectives of assessment. Fitness for purpose is the key determinant: if teachers and school administrators are only aiming for an informal evaluation

of students' progress, then existing informal assessment practices may be sufficient; if however, they would like to identify weaker students for enrichment and support, for example, then more accurate and reliable instruments may be useful. It is important to emphasize that contrary to common belief, the use of oral tests does not imply the push for high-stakes counterproductive literacy testing in early childhood, but rather they contribute to a 'balanced' mixed-method approach to assessment in general (Dockrell, 2001), which should be encouraged.

While the development of new instruments specifically for young L2 or bilingual learners should be welcomed, it should be recognized that existing and widely used L1 measures, such as the ones sampled in this study, do provide an excellent basis, at the very least, for future work in this field.

References

- AERA. (1999). *Standards for Educational & Psychological Testing*. Washington, D.C.: American Psychological Association, and National Council on Measurement in Education.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Barona, A., & Santos de Barona, M. (2000). Assessing Multicultural Preschool Children. In B. A. Bracken (Ed.), *The Psychoeducational Assessment of Preschool Children*. Boston: Allyn & Bacon.
- Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, Biliteracy, and Learning to Read: Interactions Among Languages and Writing Systems. *Scientific Studies of Reading*, 9(1), 43-61.
- Blatchford, P., Burke, J., Farquhar, C., Plewis, I., & Tizard, B. (1987). Associations between preschool reading related skills and later reading achievement. *British Journal of Educational Psychology*, 13, 15-23.
- Bracken, B. A. (1994). Advocating for effective preschool assessment practices. A comment on Bagnato and Neisworth. *School Psychology Quarterly*, 9(2), 103-108.
- Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read: A causal connection. *Nature*, 301, 419-421.
- Bryant, P., & Bradley, L. (1985). *Children's Reading Problems*. Oxford: Blackwell.
- Bryant, P., Maclean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology*, 26, 429-438.
- Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical instrumentation. *Language, Speech, and Hearing Services in Schools*, 32, 38-50.
- CDC. (2006). *Guide to the Pre-primary Curriculum*. Retrieved from.
- Chow, B., & McBride-Chang, C. (2003). Promoting Language and Literacy Development through Parent-Child Reading in Hong Kong Preschoolers. *Early Education and Development*, 14(2), 233-248.

- Clay, M. (1972). *The Early Detection of Reading Difficulties: A Diagnostic Survey with Recovery Procedures*. Auckland: Heinemann Education Books.
- Dockrell, J. E. (2001). Assessing Language Skills in Preschool Children. *Child Psychology & Psychiatry Review*, 6(2), 74-85.
- Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British Picture Vocabulary Scale (Second Edition) Manual*: nferNelson Publishing Company Ltd.
- Evangelou, M., Brooks, G., Smith, S., & Jennings, D. (2005). *A longitudinal evaluation of the Peers Early Education Partnership (PEEP) 1998-2005*. London: DfES.
- Garber, H. L., & Slater, M. (1983). Assessment of the culturally different preschooler. In K. D. Paget & B. A. Bracken (Eds.), *The psychoeducational assessment of preschool children* (pp. 443-471). New York: Grune & Stratton.
- Johnston, P. H., & Rogers, R. (2003). Early Literacy Development: The Case for "Informed Assessment". In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of Early Literacy Research* (pp. 377-389). New York: The Guilford Press.
- Johnston, R. S., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in pre-readers: The nature of the relationship. *Reading and Writing: An Interdisciplinary Journal*, 8, 217-234.
- Justice, L. M., Invernizzi, M. A., & Meier, J. D. (2002). Designing and Implementing an Early Literacy Screening Protocol: Suggestions for the Speech-Language Pathologist. *Language, Speech, and Hearing Services in Schools*, 33, 84-101.
- Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year old children. *Journal of Educational Psychology*, 90, 294-311.
- McLaughlin, B., Gesi Blanchard, A., & Osanai, Y. (1995). *Assessing language development in bilingual preschool children*. Washington, D.C.: National Clearinghouse for Bilingual Education.
- Nagle, R. J. (2000). Issues in Preschool Assessment. In B. A. Bracken (Ed.), *The Psychoeducational Assessment of Preschool Children*. Boston: Allyn & Bacon.
- Salinger, T. (2003). Assessing the Literacy of Young Children: The Case for Multiple Forms of Evidence. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of Early Literacy Research*. New York: The Guilford Press.
- Salvia, J., Ysseldyke, J. E., & Bolt, S. (2006). *Assessment in Special and Inclusive Education (Tenth Edition)*. Boston: Houghton Mifflin Company.
- Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. New York: Guilford Press.
- Tong, N., & Clem, W. (2006, May 13, 2006). Kindergarten drilling tactics outdated: EMB (Pre-school sector comes in for stern criticism in report on standards of teaching). *South China Morning Post*, p. E1.
- Washington, J. A., & Craig, H. K. (1999). Performances of At-Risk, African American Preschoolers on the Peabody Picture Vocabulary Test-III. *Language, Speech, and Hearing Services in Schools*, 30, 75-82.

- Whitehurst, G. J., & Lonigan, C. J. (1998). Child Development and Emergent Literacy. *Child Development*, 69(3), 848-872.
- Yaden, D., Rueda, D., Tsai, T., & Esquinca, A. (2004). Issues in Early Childhood Education for English Learners: Language Proficiency Assessment, Professional Development and the Outcomes of Early Literacy Intervention. In O. N. Saracho, and Spodek, B. (Ed.), *Contemporary Perspectives on Language Policy and Literacy Instruction in Early Childhood Education* (pp. 215-241): Information Age Publishing.