International Journal of Learning Management Systems

http://dx.doi.org/10.12785/ijlms/020101

Development of a Diagnostic Instrument to Identify Reading, Writing and Numeracy Difficulties in Arabic before Entry to Primary School in Bahrain

Amina Alharmesi Alhajeri [1,*] and Jim Boyle[2,*]

[1]*Psychology Department, College of Arts, University of Bahrain, Bahrain*

[2]*Director of Postgraduate Professional Training in Educational Psychology, School of Psychological Sciences and Health, University of Strathclyde, 40 George Street, Glasgow G1 1QE, UK*

Email: harmas2@yahoo.com, j.boyle@strath.ac.uk

Received: *3 July 2013; Revised: 2 Aug. 2013; Accepted: 16 Sep. 2013*
Published online: 1 Jan. 2014

Abstract: This study aimed at developing a reliable instrument to identify reading, writing and numeracy problems at school entry in Bahrain. Three curricular areas, reading, writing and numeracy were targeted for building this indtrument. The reading section consisted of items derived from the research literature relating to phonological awareness and concept of print. The writing section consisted of items requiring the child to write their own name, draw a human figure, write a letter to a favourite person and copy geometric shapes. The numeracy section composed of addition items derived from the Maths Recovery programme (Wright, Martland, & Stafford, 2006). Item analyses and analyses of reliability and validity were carried out.

A sample consisting of 145 Bahraini children from preschool (N=47, 14 boys and 30 girls, mean age 72.77 months, range 63 months to 95 months) and grade one (N=98, 47 boys and 51 girls, mean age 79.63 months, range 71 months to 115 months) were used for validation. All of the preschool children had teacher-identified learning difficulties and one group of the grade one pupils (N=48, 22 boys and 26 girls, mean age 82.33 months, range 76 months to 115 months) had learning difficulties while a second group (N=50, 25 boys and 25 girls, mean age 76.92 months, range 71 months to 84 months) were typically-developing.

The results revealed satisfactory levels of internal consistency reliability for the final selection of items (Cronbach alpha was 0.73) and concurrent validity based upon teacher evaluation was in the range 0.81-0.90. The instrument also distinguished children with difficulties from their typically-developing peers (insert some details). Analyses of sensitivity and specificity based upon an additive risk model (reference) yielded sensitivity of 0.93 and specificity of 0.67, in both cases based upon problems in two or more of the three curricular areas of the test.

The instrument reliably identified over 90% of the children with difficulties and can serve as a diagnostic tool to assist teachers in assessing problems in literacy and numeracy. However, in the light of the lower levels of specificity, use of the instrument as a screening test is not recommended.

Keywords: diagnostic instrument, reading difficulties, writing difficulties numeracy, entry to primary preschool.

Introduction

Screening tests are commonly used by teachers in primary school in order to identify children at risk of academic difficulties (Howell & Nolet, 1999 in Scott, Hintze & Floyed, 2008). The focus of such screening tests is on basic skills in social behaviour, speech, language, early literacy and numeracy. An efficient screening test is judged by its sensitivity, the ability to identify children truly at risk of failure, and specificity which is its ability to identify children truly not at risk of failure. An effective screening test should have sensitivity and specificity of at least 80% (Jenkins Hudson, & Johnson, 2007; Butler, 1988). Table 1 shows some examples of screening tests which were used with English speaking children either in preschool or grade one in the United States. It shows that most of the tests (*) have a sensitivity level 80 % and above, but the specificity was below the satisfactory level in most of them. One reason for that is that the test developers or examiners tend to select a low cut-off score in order to identify as many children truly at risk as possible. For instance, Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test was used in three studies and the researchers used different cut-off scores. Johnson, Jenkins, Petscher and Catts (2009) selected a cut-off score that gave a sensitivity level of 90%. This procedure reduces the level of the test specificity which represents the children with no problem. However, Texas Primary Reading Inventory (TPRI) was designed to identify those children who are not at risk for later reading failure. They developers made more emphasis on the importance of false negative errors than false positive errors in identifying reading difficulties. The way of administration might also influence the sensitivity of the test. For instance, when a test is administered to each child individually, it allows the examiner to observe the child and identify some important features of his learning which might be difficult to find under group administration. As a result more children with difficulties will be identified and this increases the sensitivity of the test. A sample size is also important for test validity. A large sample size could be more representative of the population and it increases the chance of identifying more children with difficulties than a smaller sample size. The quality of the reference test and the rational of using a cut-off score also have an impact on the sensitivity of the test Nelson, 2009). The timing of the test might have an effect on the rate of sensitivity. For instance, for improving the accuracy of (DIBELS), it was suggested to delay the first administration timing of each subtest in order to find the most optimal time. This procedure might lead to higher accuracy in Letter Name Fluency (LNF), Nonsense Word Fluency (NWF), and Oral Reading Fluency (ORF). Also it was suggested to use the Initial Sound Fluency (ISF) assessments as a “first cut” to identify children who do not meet the pass criteria and use another phonological awareness measure in order to reduce the false-positives.

Table 1

Predictive Validity for Examples Screening Tests in Reading, Writing and Maths (for Preschool and Grade 1) during the Period 1990-2010

Study	Age/level	Sample size	Administration	Test	Sensitivity	Specificity	Reference Test
Diamond (1990)	58.75 months	92 children All children in the preschool	Individual	Revised Denver Developmental screening tests (RDDST) Maths Reading	0.38 0.67	0.79 0.86	SAT Achievement Test Scores (2,3, 4 grades)
Flynn & Rahbar (1998)	Preschool	708 children Group	Individual	Literacy Screening Battery (LSB)	0.80*	0.72	The Peabody Picture Vocabulary Test-III (PPVT-R (Dunn and Dunn, 1981)
Erford, Dorman, Ivey & Wingeart, (2001)	Grade 1	449 children from two schools in central Maryland	Group	Writing Essential Skill Screener - Elementary Version (WESS-E)	0.92*	0.76	Woodcock Johnson: Tests of Achievement -Revised (WJ-R) Broad Written Language (Woodcock & Johnson, 1989, 1990)
VanDerHeyden, Witt, & Naquin (2003)		282 children part of school wide screening	Small group	Problem Validation Screening (PVS) Maths Reading	0.71 *	0.89 0.99	Criterion Assessment Woodcock , Johnson Psycho educational Battery-Revised (W-R); The Iowa Test of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1993)
Hintze, Ryan, & Stoner, (2003)	Preschool	86 children from three elementary schools and were from 10 different kindergarten classrooms.	Individual	Dynamic Indicators of Basic Early Literacy Skills (DIBELS) cut off score 15 to 25 LNF	0.80* 0.91*	0.72 0.39	Comprehensive Test of Phonological Processing (CTOPP)
Nelson (2008)	Preschool	177 children from 10 classrooms from 2 different schools	Small group of two to five	PSF ISF LNF PSF NWF	0.86-0.89* 0.91-0.94* 0.53-0.72 0.89-0.94* 0.82-0.94*	0.53-0.61 0.30-0.27 0.70-0.69 0.42-0.38 0.59-0.53	The Test of Phonological Awareness – Second Edition: Plus (TOPA-2_; Torgesen & Bryant, 2004a) Cut-off score 90. The Woodcock–Johnson Tests of Achievement, Third Edition (WJ III; Woodcock, McGrew, & Mather, 2001) in small group and cut-off score 85
Johnson, Jenkins, Petscher and Catts, (2009)	Grade 1 and 2	12,055 children from 309 schools in 33 districts across Florida State public schools.	Individual and small group	NWF LNF	0.90* 0.90*	0.34 0.45	Florida Comprehensive Assessment Test—Sunshine State Standards (FCAT)+ the SAT Reading Comprehension subtest for Grade

Watkins and Edwards (2004).	First grade	1204 All children	Individual	Mountain Shadows Phonemic Awareness Scale (MS-PAS).	0.94*	0.68	Gates-MacGinitie Reading Test (MacGinitie and MacGinitie, 1989)
Bradley, Erford and Stephens (2005)	6-8 (Grade 1)	171 children from four schools in central Maryland.	Individual	The Reading Essential Skills Screener-Elementary Version (RESS-E):	0.94	0.86	Teacher judgment
Nelson (2008)	Preschool (Mean age 5.44)	180 children selected from 10 classrooms from 2 different schools.	Individual	Texas Primary Reading Inventory (TPRI) Graphophonemic Knowledge Mid-Year(TPRI-GK Mid) Texas Primary Reading Inventory Phonemic Awareness Mid-Year(TPRI-PA Mid)	0.78 0.78	0.78 0.43	Woodcock-Johnson Tests of Achievement(WJ III)
Panter,& Bracken, (2009).	Preschool	86 from two public schools in rural, western Tennessee.	Individual	The Bracken School Readiness Assessment (BSRA)	0.62	0.96	Metropolitan Readiness Tests, 6th Edition (MRT-6)
Wilson and Lonigan (2010)	42 to 55 months	176 from 21 preschools in north Florida	Individual	GRTR-R (Whitehurst, 2001) and IGDIs) Early literacy GRTR-R IGDIs score Print Knowledge GRTR-R IGDIs score Definitional Vocabulary GRTR-R IGDIs score Phonological Awareness GRTR-R IGDIs score	0.90* 0.93* 0.92* 0.94* 0.95* 0.95* 0.93* 0.93*	0.69 0.38 0.56 0.40 0.15 0.06 0.23 0.13	TOPEL (Lonigan et al., 2007).
Fuchs, Fuchs, Compton, Bryant, Hamlett and Seethler (2007)	Grade one	225 children from 41 classrooms In 10 schools	Large group	The screening measures for forecasting math disability (MD)	0.69	0.79	Woodcock-Johnson III (WJ III; Woodcock, McGrew, & Mather, 2001)
Geary, Bailey, and Hoard, (2009).	Preschool	228 children from 12 elementary Schools	Group	Preschool Number Sets Test	0.69	0.67	Numerical Operations and Word Reading subtests from the Wechsler Individual Achievement Test-II-Abbreviated (Wechsler, 2001a, 2001b).

Note: Letter Name Fluency (LNF), Initial Sound Fluency (ISF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), and Oral Reading Fluency (ORF)

Phonological awareness at rhyme level at an early age is a strong predictor for successful reading in the future (Wagner & Torgeson, 1987; Blachman, 1984; Byrne & Fielding-Barnsley, 1991; Bryant, Bradley, Maclean, & Crossland, 1989; Parrila, Kirby, & McQuarrie, 2004; Boscardin, Bengt, Francis & Baker, 2008). Research found that children at 3- 4 years learn rhyme (Bradley & Bryant, 1983; Maclean, Bryant & Bradley, 1987; Pullen & Justice, 2003). Hatcher, Hulme and Snowling (2004) found in a longitudinal interventional research that typically developed children age 4-5 years had sufficient phonological awareness, and did not need addition training in phonological awareness. It was found that phonological awareness is related to failure in both reading and writing (Bradley & Bryant, 1983; Vellutino & Scanlon, 1987). Phonological awareness helps children to focus on the grapheme and link it to the phonemes (Adams, 1990; Bryant, et al., 1989; Byrne & Fielding-Barnsley, 1991; Tunmer, Herriman, & Nesdale, 1988).

Johns (1982) indicated that concepts of print are a consequence as well as a cause of reading progress. Dickinson, McCabe, Anastopoulos, Peisner-Feinberg, and Poe (2003) found a bidirectional relationship between phonological awareness and print knowledge, at least until the end of kindergarten. They showed that both vocabulary and phonological awareness predict print knowledge. Nichols, Rupley, Rickelman, & Algozzine (2004) reported that print is important for reading. Hammill, (2004) and Snow, Burns, & Griffin (1998) revealed that concepts of print are among the best predictors of later reading. Other researchers reached similar findings (Bowey, 1994; Burgess & Lonigan, 1998; Ehri, 1998; Johnston, Anderson, & Holligan, 1996; Lonigan, Burgess, & Anthony, 2000; Stahl & Murray, 1994; 1997; Stuart, 1995; Badian, 2001).

Riley (1995) found a correlation between the children's letter naming and name writing when they enter school, and their reading at the end of the grade 1. He indicated that both letter naming and name writing were strong predictors for reading. In another study, Riley (1996) found a relationship between name writing, and word, and non-word reading. Bloodgood (1999) found an association between recognition of name and age in the 3 years old children; whereas name production had a relationship with alphabet knowledge, word recognition, and concept of word for 4 - 5 years old. She indicated that the quality of the name writing could provide an idea about the quality of motor control. Behnken and Haney (2002)supported Bloodgood findings. They confirmed that name writing was associated with word reading only. They added that name writing is important for reading prediction skills. It provides information such as development of dictation, spelling, word recognition, and concept of word. It was revealed that the skills which are related to print such as knowledge of letters, name writing, and concepts of print are the best predictor for future reading (Hammill, 2004; Snow, Burns, & Griffin, 1998).

Haney and Behnken, (2002) found that name writing could be used to identify the children at risk of developing reading problems by checking phonological awareness, rapid naming, and letter knowledge (Adams, 1990; Perfetti, Beck, Bell, & Hughes, 1987; Torgesen, Wagner, Rashotte, 1994; Vallutino & Scanlon, 1987) to predict future reading. They mentioned that using name writing in a reading screening test might reduce the false positive cases. They found that a relationship between name writing, and alphabet knowledge, and concepts of print. They concluded that name writing could be used as a predictor for writing too. Dunsmuiri and Blatchford (2004) found a relationship between a child's ability to write his own name at school entry and his writing outcome later.

Weil and Amundson (1994) found an association between copying oblique and writing letters. They mentioned that it can be used to identify the children at risk of handwriting problems. Marr and Cermak (2001) observed that oblique lines predicted handwriting in grade 1.
Mti-zissi, Zafiropoulou and Bonti (1998) found that children's drawing of human figure could be a useful diagnostic tool for identifying children at risk of having dyslexia before they enter primary school. Mti-zissi and Zafiropoulou (2001) found that children's drawing of human figure at preschool could be used to identify children with learning problems in future. Bonoti, Vlachos, and Metallidou (2005) found a positive correlation between the scores of the children on drawing of human figure and writing tasks. Re, Caeran, and Cornoldi, (2008) found that writing a letter is useful in identifying children with writing problems.

Secada, Fuson and Hall (1983) found that counting on strategy in addition depends on counting forward from a certain point on the number word sequence. Ho and Fusson (1998) observed that forward number word sequence is important for a simple addition task and understanding counting in teens. Okamoto and Case (1996) found that the link between counting and quantity discrimination in preschool predicts future achievement in maths (Gersten, Jordan, & Flojo, 2005). They claimed that these abilities are not given enough attention in preschool. Counting is another strong precursor for future maths (Aunola, Leskinen, Lerkkanen, & Nurmi, 2004). Johansson (2005) found that a child could use counting on the number word sequence at an early stage. As the child gets more experience in counting, he might discover a form of regularities in the number word sequence which he could use to develop new correct strategies to solve some arithmetic problems.

In Bahrain and other Arabian Gulf countries there is no reliable and valid screening or diagnostic instrument to identify young children learning difficulties before they join primary school. At the same time, in Bahrain there is an increased percentage of failure and withdraw from school. It is possible that absence of screening or diagnostic tests and inappropriate support service contributes in these problems. It is inappropriate to use a screening test from other Arabic countries in the Middle East because they have their own local Arabic and the classical Arabic is not in use even in the class. Another point is Bahrain has limited financial and expertise resources which make it unwise to have screening tests for each problem or incidence. Therefore, the aim of this study is to develop a reliable and valid test confirmatory screening to identify young Bahraini learning difficulties in reading, writing and numeracy before they enter school.

Method

Participants

The number of children who participated in this research was 145 Bahraini children from preschool (N=47, 14 boys and 30 girls, mean age 72.77 months, range 63 months to 95 months) and grade one (N=98, 47 boys and 51 girls, mean age 79.63 months, range 71 months to 115 months). All of the preschool children had teacher-identified learning difficulties and one group of the grade one pupils (N=48, 22 boys and 26 girls, mean age 82.33 months, range 76 months to 115 months) had learning difficulties while a second group (N=50, 25 boys and 25 girls, mean age 76.92 months, range 71 months to 84 months) were typically-developing. Informed consent was collected from the children's parents.

Materials and Procedure

The following tasks were used:

1. Pictures of familiar objects. Each picture was placed in front of a child. The researcher says the name of the object in the picture and the child is asked to identify the first sound he hears in the word. The number of the tasks was 17
2. Three pictures are placed in front of a child. The researcher says the name of the object in each picture one at a time and the child is asked to identify the two words that begin with the same sound. 5 tasks were administered.
3. Three pictures are placed in front of a child. The researcher says the name of the object in each picture one at a time and the child is asked to identify the two words that end with the same sound. 5 tasks were administered also
4. Blank papers and pencils were used. Writing ones' own name, drawing of human figure, writing a personal letter and copy four shapes were used to identify children with writing problems. Children performance in writing at the school is collected at the end of the year. The child is asked to do the following: "I want you to draw a person that you like the most". If he finishes the drawing, the researcher says: "Now I want you to write a letter to this person". A coding scheme was developed for the drawing, writing the letter, and name of the child as part of this research.
5. Four plastic geometrical shapes were placed one at a time in front of the child and he was asked to copy on the blank paper provided. These shapes are a circle, square, triangle and a diamond. The researcher says: " I want you to look carefully at the …. and then draw it on the paper".
6. In numeracy additive tasks borrowed from Maths recovery were used. Counters and two screens were used to present the tasks to the child. At the beginning, the researcher placed in front of the child two screened collections of counters 3 and 1 and then 5 and 4. The child was asked to say how many counters altogether. If his answers were correct the researcher placed 9 and 6. If the child was able to do the task, the assessment ends at this point. If the child could not do the first two tasks, the researcher presented 3 tasks in which the first collection was screened and the second was visible. These tasks were 5and 2 and 7 and 5. If the child was able to solve the tasks correctly, the assessment ends. If the child experienced difficulties with the tasks, the same tasks were repeated, but with no screen. If the child still had problem with the tasks, the researcher placed in front of the child a collection of 13 counters in a line and asked the child to count them. Then, a collection of 18 counters were placed and the child was asked to do the same (*Wright et al, 2006*).

Results

Item analysis using item - Total correlation

Item analysis is used to determine the internal consistency of the test.

Table 2
Item Analysis for the Items in the Test (n=95)

Item	Scale Mean if Item deleted	Scale Variance if Item deleted	if corrected item- Total Correlation	Cronbach's Alpha if item deleted	Item diff.
1. Identify the initial sound a word	3.95	4.42	0.46	0.69	0.64
2. Identify the shared sound in the beginning of 2 words	4.52	5.32	0.25	0.72	0.07
3. Identifying the shared sound at the end of 2 words	4.54	5.29	0.34	0.72	0.05
4. Identifying the front part of the storybook	4.43	4.91	0.40	0.70	0.16
5. Identifying the back part of the storybook	4.44	4.95	0.39	0.71	0.15
6. Identifying some letters	4.00	4.50	0.45	0.70	0.59
7. Identifying some words	4.13	4.73	0.33	0.72	0.46
8. The child's strategy in the arithmetic task	4.48	5.21	0.27	0.72	0.11
9. Human drawing of a favourite person	3.80	4.91	0.31	0.72	0.76
10. Copying shapes	4.45	5.19	0.25	0.72	0.14
11. Writing own Name	3.80	4.79	0.40	0.70	0.78
12. Writing a letter	3.91	4.49	0.50	0.69	0.68

The results in Table 2 are for group 1. Group 1 is the group of both preschool and grade 1 children with learning problems. Cronbach's Alpha was used to determine the internal consistency among the items in the test. Test reliability was maximal at 0.73, when the remaining three items holding the storybook, and opening it, identifying the story in the book, and beginning reading from right to left were removed from the test. Cronbach's Alpha was used in order to show how well the items in the test measure a single one-dimensional latent construct. If Cronbach's Alpha value is above the satisfactory level 0.70, it means that the items measure the same construct and the test is reliable. Column 3 in the table shows the point biserial correlation. This correlation is appropriate for my data in concepts of print section were nominal, and the children's outcomes on competencies of reading, writing, and numeracy were also dichotomous (yes/no). The point biserial correlation determines the quality of the items in the test that is the extent to which they measure the same construct. The point biserial values for most of the items in column 3 were above the satisfactory level of intended consistency reliability 0.25. The

results show that the items in the test were of good quality (Varma, 2006; Rust, & Golombok, 1989; Hogen, 2007).

Column 5 displayed item difficulty. It refers to the percentage of people who answer an item in the test correct. In most tests item difficulty is in the range 0.30-0.70 (Kaplan & Saccuzo, 2005; Howitt & Cramer, 1997; Hogen, 2007). Item 3 is very difficult for the children; but it was of good quality, and removal of this item resulted in a decrease in Cronbach's Alpha value. It means it measures the same construct as other items.
Item difficulty was calculated to explore how difficult the items were for group (1). In column 5, one could observe two criteria. First, the values were distributed across the whole range 0 to 1. Second, there was more concentration towards the centre. These two criteria increased the reliability of the test (Varma, 2006). Although the children experienced difficulties on items 4, 5, 8, and 10, these items had good quality, and the quality of the items was more important than the items' level of difficulty.

Validity for the test

A test is valid if it measures what it is intended to measure (Domino & Domino, 2006; Rust, & Golombok, 1989). For this test, analyses of criterion related validity was run. In criterion related validity, both analyses of predictive and concurrent validity were used (Rust, & Golombok, 1989; Cronbach 1984, Hogen, 2007). In addition, construct validity was used.

Criterion related validity

This part includes a description on the validity of the test using the predictive and concurrent validity.

Predictive validity

The additive risk model is a simple and practical method to investigate areas of problems in children's performances without using statistical software. Before going into details about this procedure, finding the predictive validity for the test, and investigating the risk factors; it is important to define both sensitivity and specificity. Sensitivity is the test capacity to identify children with a problem such as learning problems in this research (Butler, 1988). The higher the sensitivity the more children with learning problems will be correctly identified and the fewer the false negative cases. Specificity is the capacity of the test to identify children without a problem (Butler, 1988). The higher the specificity, the more typically developed children were correctly identified, and the fewer the false positive cases. This procedure is useful in identifying both false positive and false negative cases. In the additive risk procedure I used the data for grade 1 children in both group (1) and (2) only because I had data for these two groups.

The present test consists of four areas:

- Phonological awareness
- Concepts of print
- Writing part
- Numeracy

Adding to that, I used the "school outcomes" in reading, writing and maths. These are teacher's evaluation of the children at the end of the year.

Table 3

Predictive Validity and Factors for Future Persistence Learning Problems in Reading (n=98)

No. of children with no Progress /	No. of children with progress	Area of the test
Difficulties in any one or more areas	17	48
No Difficulties on the test	0	33
Sensitivity = 14/17 = 82.4 %		
Specificity = 33/81 = 40.70 %		
Difficulties in any two or more areas	14	27
No difficulties or difficulties in any one area only	1	54
Sensitivity = 14/15 = 93.3 %		
Specificity = 54/81 = 66.6 %		
Difficulties in any three or more areas	7	9
No difficulties or difficulties in less than three areas of the test	5	68
Sensitivity = 7/12 = 58.3 %		
Specificity = 68/ 77 = 88.3 %		
Difficulties on all four areas	3	7
No difficulties or difficulties in three or fewer areas	13	79
Sensitivity = 3 / 16 = 18.8 %		
Specificity = 79 / 86 = 91.9 %		

The results in Table 3 illustrate using a criterion of problems in two areas rather than just one optimized sensitivity and specificity for reading. The acceptable value for both sensitivity and specificity in a screening test is not less than 80% (Butler, 1988). In the present test, the optimal level of sensitivity was 93.3%, when the difficulty was in two areas of the test or more. This value was extremely high. It means that 93.3% of the children with difficulty in two areas or more areas were identified as having difficulties in reading. It means the test could not identify 6.7 % of the false negative cases. The specificity in the same areas of the test was only 66.6%, which was below the acceptable standard. It means 66.6 % of the children were identified with no problems in reading. In this case the false positive cases are 33.4%. It shows that the false negative cases were few and this increases the test validity. Because the value of specificity was below 80%, the test is more useful as a diagnostic test than a screening test.

Table 4
Predictive Validity and Factors for Future Persistence Learning Problems in Writing and Numeracy Test (n=98)

	No. of children with no progress /	No. of children with progress
Difficulties in any one or more areas	26	44
No Difficulties on the test	0	28
Sensitivity = /16 = 100%		
Specificity = 27/71 = 38%		
Difficulties in any two or more areas	14	30
No difficulties or difficulties in one area only	2	59
Sensitivity = 14/16 = 87.5 %		
Specificity = 59/89 = 66.3 %		
Difficulties in any three or more areas	6	24
No difficulties or difficulties in less than three areas of the test	8	61
Sensitivity = 6/14 = 42.9 %		
Specificity = 61/ 85 = 71.8 %		
Difficulties on all four areas	4	7
No difficulties or difficulties in three or fewer areas	12	75
Sensitivity = 4 / 16 = 25%		
Specificity = 75 / 82 = 91.5 %		

The results in Table 4 show similar results for the difficulties in both writing and numeracy. In the present test, the optimal level of sensitivity was 87.5% when the difficulty was in two areas of the test or more. This value was above the acceptable value in a screening test. It means 87.5% of the children with difficulties in two areas or more were identified as having problems in either writing or numeracy. The false negative cases were 12.5 %. The specificity in the same areas of the test was only 66.3%. It is below the acceptable standard. It means that only 66.3 % of the children were identified with no problems in writing or numeracy. The false positive cases were 33.7%. If we compare it with the reading results, the test is better in identifying more false negative cases in reading than either in writing and numeracy.

Concurrent validity

Concurrent validity is a correlation between scores from two related tests (Rust & Golombok, 1989; Hogen, 2007). Test scores of all the children, and the teacher's evaluation of the children's achievement were used to determine the concurrent validity.
Point biserial correlation was used for phonological awareness tests, because the scores were interval and the teacher judgment was dichotomous.

Table 5
Point Biserial Correlations between the Children Scores on the Test and Teacher Evaluation of the Children's Attainment (n=145)

Item	1	2	3
Initial sound(phoneme) in a word			
The shared sound (phoneme)in the beginning of two words	0.81**	___	
The shared sound at the end of two words	0.83**	.90**	____

* Correlation is significant at .05 level (2 tailed)
** Correlation is significant at .01 level (2 tailed)

The results in Table 5 show the correlation value for all items is in excess of r = 0.81. This means there is a high agreement between the class teacher's evaluations of the children academically and the children's scores on the test. The results indicate a high level of concurrent validity. It is concluded that the test achieved satisfactory levels of validity.

Construct validity

Construct validity was used to examine if the test measures children's problems in reading, writing and numeracy (Salvia & Ysseldyke, 2006). Point biserial correlation was run to find out if the test discriminates between the two groups' achievements on the three tasks of phonological awareness. The effect size was used to determine the strength of the items, and chi square tests to explore the extent to which the test discriminates children with learning problems from typically developed children. The chi square tests were more appropriate for concepts of print, writing, and numeracy items; because the scores were nominal, and ordinal.

Table 6 shows that the mean scores of the typically developed children (group 2) on the three items of phonological awareness were higher that the means scores of the children with learning problems (group 1). The t-tests show significant differences between the typically developed children and the children with learning problems on the three items of phonological awareness. It means that the scores of the three items of phonological awareness identified the children with learning problems.

Table 6.
Comparison between the Children with Learning Problems Group (1) and Typically Developed Children Group (2) on Phonological Awareness Variables (n=145)

Item	Group	No	Mean score	Std.	t
Initial sound (phoneme)in a word	with problems	95	6.01	3.94	(141.20)= -21.48**
	Typ. developed	50	16.30	1.82	
The shared sound (phoneme) in the beginning of two words	with problems	95	0.69	1.06	(141.97) = - 29**
	Typ. developed	50	4.72	0.61	
The shared sound at the end of two words	with problems	95	0.44	0.92	(143) = - 4.12**
	Typ. developed	50	4.56	0.70	

* Correlation is significant at .05 level (2 tailed)
** Correlation is significant at .01 level (2 tailed)

Effect size is a measure of the strength of the relationship between two variables (Howitt & Cramer, 1997). I determined the effect size for each item in order to find the strength of each item in discriminating between the children in the two groups in reading. The differences between the means score of each group and the standard deviation was used to calculate the effect size. The calculated effect sizes for the three items are as follows:

Table 7
The Strength of the Effect Size for Phonological Awareness Items

Item	Effect size(d)	Confidence interval
Initial sound (phoneme)in a word	3.04	2.55 - 3.53
The shared sound (phoneme) in the beginning of two words	4.29	3.69 - 4.90
The shared sound at the end of two words	4.81	4.16 - 5.46

In Table 7 the effect size shows that identifying the initial phoneme in a word is poorer at discriminating between the two groups than identifying the shared sound (phoneme) in the beginning of two words, and identifying the shared sound at the end of two words. The last two items are equally discriminating.

The data for concepts of print were nominal (yes/no) and for writing and numeracy ordinal so chi square tests were an appropriate choice to investigate the differences between groups (1) and (2) on these areas.

Table 8.
Comparison between Children with Learning Problems (group 1) and Typically Developed Children (group 2) on Concepts of Print, Writing and Numeracy (n=145)

Item	X^2
identifying front of the book	37.81**
Identifying back of the book	66.30**
identify letters	27.72**
Identify words	39.61**
writing own name	10.26**
Drawing human figure	25.21**
writing a letter	48.17**
copying shapes	67.26**
Numeracy strategy	108.69**

* Correlation is significant at .05 level (2 tailed)
** Correlation is significant at .01 level (2 tailed)

The results in Table 8 show that group (2), the typically developed children, performed better than group (1) children with learning problems on all items in the three areas. This means that the test scores can detect reading, writing and numeracy problems in young children. Identifying the back part of the storybook was the most discriminating item of concepts of print. It can be used to discriminate children with difficulties in reading from children with no difficulties in reading (Lipsey & Wilson, 2001). Copying a diamond (shapes) was the most discriminating in writing between the typically developed children

(group2) and the children with learning problems (group1). The counting strategy in numeracy also discriminated between the children in the two groups in numeracy. It can be concluded that the test achieved acceptable levels of validity.

Discussion

In Bahrain, an Arabic gold standard test is not available to identify young children's problems in literacy or numeracy. This research is an attempt to develop in Arabic a reliable and valid test to locate Bahraini children who are likely to have learning problems in reading, writing, and numeracy. For the test reliability, the highest Cronbach Alpha value was 0.73 when 12 items of the test were selected. This value is above the satisfactory value which is 0.70 (Varma, 2005).

For the validity, predictive, validity the additive risk model (Whitehurst & Fischel, 1994) shows that two or more areas of the test predict future reading problems in reading. The four areas were phonological awareness, concepts of print, writing and numeracy. The value of the sensitivity index was 93.3% and that for specificity was 66.6 %. The percentage of false positive cases where children had no problems and the test identified them as having problems was high. In writing and numeracy two areas of the test or more also predicted future problems in writing and numeracy. The sensitivity was 87.5% which is above the satisfactory level; but the specificity was 66.3% which is less than 80%. The problematic level of specificity limits the utility of the test as a universal population-screening instrument, which could be administered to all children to predict subsequent problems in reading, writing and numeracy as the number of false positives would be unacceptably high. Boyle, Gillham and Smith (1996) had low specificity in their study. They indicated that, in this case, the test is more appropriate as a competence based diagnostic instrument, and may also be valuable as a confirmatory screening test, which can be used with suspected cases of having problems in literacy and numeracy. The test provides a profile of children's strengths and weaknesses in reading, writing and numeracy. Class teachers could use the profile to help children, whom they suspect to have learning problems. Application of the test as a diagnostic test could reduce false positive cases. It can be concluded that the test is reliable and valid.

The concurrent validity showed high agreement between the teacher evaluation of the children's performances in the class and the test results of the children ($r = 0.83$).
In the construct validity, the test was able to discriminate between the children with learning problems and the typically developed children on phonological awareness, concepts of print, writing one's own name, drawing a human figure, writing a letter and copying shapes. It can be concluded that the test is reliable and valid as a diagnostic test to identify Bahraini children with reading, writing and numeracy difficulties. Based on the findings, the teacher needs to consider the level of attainment of each child when she plans her lesson.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge: MIT Press.

Aunola, K., Nurmi, J., Niemi, P., Lerkkanen, M., & Rasku-Puttonen, H. (2002). Developmental dynamics of achievement strategies, reading performance, and parental beliefs. In Georgiou, G., Manolitsis, G., Nurmi, J., & Parrila, R. (2010). Does task-focused versus task-avoidance behavior matter for literacy development in an orthographically consistent language? *Contemporary Educational Psychology 35, 1–10.*

Badian, N. A. (2001). Phonological and orthographic processing: Their roles in reading prediction. *Annals of Dyslexia, 51*, 179-202.

Bloodgood, J. W. (1999). What's in a name? Children's name writing and literacy acquisition. *Reading Research Quarterly, 34*(3), 342–367.

Bonita, F., Vlachos, F., & Metalloid, P. (2005). Writing and drawing performance of school age children. Is there any relationship? *School Psychology International 26*(20), 243-255.

Bowey, J. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. In Justice, L. Bowles, R. & Skibbe, L. (2006). Measuring preschool attainment of print-concept knowledge: A study of typical and at-risk 3- to 5-year-old children using item response theory. *Language, speech and hearing services in schools, 37* 224-235.

Boyle, J., Gillham, B. & Smith, N. (1996). Screening for early language delay in the 18-36 month age-range: the predictive validity of tests of production, and implications for practice. *Child Language Teaching and Therapy, 2*(2), 113-127.

Bradley, L., & Bryant, P. (1983). Categorizing sounds, and learning to read: a causal connection. *Nature, 310, 419-421.*

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations between phonological awareness and reading extended to preschool letter knowledge: Evidence from a longitudinal investigation. In Justice, L. Bowles, R. & Skibbe, L. (2006). Measuring preschool attainment of Print-concept knowledge: A study of typical and at risk 3 to 5 year old children using item response theory. *Language, speech and hearing services in schools,, 37*, 224-235.

Burke, M., Hagan-Burke, S., & Kwok, O. (2009). Predictive Validity of Early Literacy Indicators From the Middle of Kindergarten to Second Grade. *The Journal of Special Education Volume 42(4) 209-226.*

Butler, J. (1988). *Child health surveliance in primary care*. London: Crown copyright.

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic Awareness to young children. *Journal of Educational Psychology, 83*(3), 451-455.

Cronbach, L. J. (1984). *Essentials of Psychological Testing* (4th ed) New York: Harper & Row.

Dickinson, D., McCabe, A., Anastopoulos, L., Peisner-Feinberg, E., & Poe, M. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95(*(3), 465–481.

Domino, G. & Domino, M. (2006). *Psychological testing: An introduction* (2nd ed).Cambridge: Cambridge University Press.

Dunsmuir, S., & Blatchford, P. (2004). Predictors of writing competence in four to seven year old children. *British Journal of Educational Psychology, 74(3),* 461-483.

Ehri, L. C. (1998). *Grapheme-phoneme knowledge is essential for learning to read words in English.* In Metsala, J. & Ehri, L. *Word recognition in beginning literacy (pp. 3-40)*. Mahwah, NJ: Erlbaum.

Erford, B. & Stephens, V. (2005).The Reading Essential Skills Screener-Elementary Version (RESS-E): Studies of Reliability and Validity. *Measurement and Evaluation in Counseling and Development 38(2), 104-14.*

Erford, B., Dorman, S., Ivey, E. & Wingeart, L.(2001). Reliability and Validity of the Writing Essential Skill Screener Elementary Version(WESS-E). *Assessment for Effective Intervention 2001; 26; 43-55.*

Fuchs, L. Fuchs, D. Compton, D. Bryant, J. Hamlett, C., & Seethaler, P.(2007). Mathematics Screening and Progress Monitoring at First Grade: Implications for Responsiveness to Intervention. *Exceptional Children 73(3) 311-30.*

Flynn, J., & Rahbar, M.(1998). Kindergarten Screening for Risk of Reading Failure. *Journal of Psycho-educational Assessment, 16; 15-35.*

Geary, D., Bailey, D. & Hoard, M.(2009). Predicting Mathematical Achievement and Mathematical Learning Disability With a Simple Screening Tool. Journal of *Psych-educational Assessment, April 7, 2009 http://online.sagepub.com*

Gersten, R., Jordan, N. & Flojo, J. (2005). Early identification and interventions for students with mathematics difficulties. In Krajewski, K., & Schneider, W. (2008). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: findings from a four year longitudinal study. *Learning instruction* (November online), 1-14.

Hammill, D. D. (2004). What we know about correlates of reading. In Justice, L. Bowles, R. & Skibbe, L. (2006). Measuring preschool attainment of print-concept knowledge: A study of typical and at-Risk 3- to 5-year-old children using item response theory. . *Language, speech and hearing services in schools, 37*, 224-235.

Haney, M. B., V., & Behnken, K. L. (2002). The relationship among name writing and early literacy skills in kindergarten. *Child study journal, 33*(2).

Hatcher, P., Hulme, C. & Snowling, M.(2004). Explicit phoneme training combined with phonemic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry, 45*(2), 338-358.

Howell, K., & Nolet, V. (1999). Curriculum based evaluation: Teaching and decision making. In Scott, A., Hintze, J., & Floyed, R. (2008). Validation and decision accuracy of early numeracy skill indicators. *School Psychology Review 37(2), 359-373.*

Howitt, D. & Cramer, D. (1997). *Introduction to statistics in psychology* (4th ed). Harlow: Prentic Hall.

Hintze, J., Ryan, A. L., & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *School Psychology Review,32, 541−556.*

Hogan, T. (2007). *Psychological testing: A practical introduction* (2nd ed). U.S: John Willey & Sons, Inc.

Jenkins J. Hudson, R., & Johnson, E. (2007). Screening for at risk readers in a response to intervention framework. *School Psychology Review 36(4), 582-600.*

Johansson, B. (2005). Number word sequence skill and arithmetic performance. *Scandinavian Journal of Psychology, 46, 157-167.*

Johns, J. L. (1982). Does our language of instruction confuse beginning readers? *Reading Psychology, 1*, 37-41.

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.

Johnson, E., Jenkins, J., Petscher, Y. & Catts, H.(2009). How Can We Improve the Accuracy of Screening Instruments? *Learning Disabilities Research & Practice, 24(4), 174–185.*

Kaplan, R. & Saccuzzo, D. (2005). *Psychological testing: Principles, applications and issues* (6th ed). U.S: Thomas and Wadsworth.

Lipisey, M., W. & Wilson, D. B. (2001). *Practical meta-analysis: Applied social research methods series*. London: SAGE Publications.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36(5),* 596-613.

MacLean, M., Bryant, P., & Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly, 33*, 255-282.

Marr, D., & Cermak, S. (2001). Consistency of handwriting development in the early elementary years: A literature review. *Israel Journal of Occupational Therapy, 10*, 109-129.

Mati-Zissi, H., Zafiropoulou, M., & Bonoti, F. (1998). Drawing performance in children with special learning difficulties. *Perceptual and Motor Skills, 87*, 487-497.

Mati-Zissi, H., & Zafiropoulou, M. (2001). Drawing performance in prediction of special learning difficulties of kindergarten children. *Perceptual and motor skills 92*, 1154–1166.

Nelson, J.(2008). Beyond correlational analysis of the Dynamic Indicators of Basic

Early Literacy Skills (DIBELS): A Classification Validity Study. *School Psychology Quarterly, 23(4)542–552.*

Nichols, W. D., Rupley, W. H., Rickelman, R. J., & Algozzine, B. (2004). Examining phonemic awareness and concepts of print patterns of kindergarten students. *Reading Research and Instruction, 43*(3), 56–81.

Okamoto, Y., & Case, R. (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. In Krajewski, K., & Schneider, W. (2008). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: findings from a four year longitudinal study. *Learning instruction* (November online), 1-14.

Okamoto, Y., & Case, R., (1996). Exploring the microstructure of children's central conceptual structures in the domain of number. In Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*, 293-304.

Panter, J. & Bracken, B. (2009). Validity of the Bracken School Readiness Assessment for Predicting first grade readiness. *Psychology in the Schools, Vol. 46(5), 397-409.*

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283–319.

Pullen, P. C., & Justice, L. M. (2003). Enhancing phonological awareness, print awareness, and oral language skills in preschool children. *Intervention in School and Clinic, 39*(2), 7-98.

Re, A. M., Caeran, M., & Cornoldi, C. (2008). Improving expressive writing skills of children rated for ADHD symptoms. *Journal of Learning Disabilities, 41*(6), 535-544.

Riley, J. (1996). *The teaching of reading: The development of literacy in the early years of school.* In Marsh, J., & Hallet, E. (1999). *Desirable literacies: approaches to language and literacy in the early years*. U.K: Paul Chapman Publishing Ltd.

Rust, J. & Golombok, S. (1989). *Modern psychometrics: the science of psychological* assessment London: Routledge.

Secada, W. G., Fuson, K. C., & Hall, J. W. (1983). The transition from counting-all to counting-on in addition. The transition from counting-all to counting-on in addition. *Journal for Research in Mathematics Education, 14*(47–57).

Stahl, S., & Murray, B. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86*(2), 221–234.

Snow, C. , Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stuart, M. (1995). Prediction and qualitative assessment of five and six year old children's reading: A longitudinal study. *British Journal of Educational Psychology, 65*, 65, 287-296.

Torgesen, J. K., Wagner, R. K., & Roshette, C. A. (1994a). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27*, 276-286.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Met linguistic abilities and beginning reading. In Lovelace, S. & Stewart, S. (2007). Increasing print awareness in preschoolers with language impairment using non-evocative print referencing. *Language, Speech and Hearing Services in Schools, 38*, 16-30.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32, 204−227.*

Varma, S. (2006). Preliminary item statistics using point biserial correlation and p value. Educational Data System, Inc. (WWW.EDDATA.COM).

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological Awareness and reading ability: Evidence from a longitudinal and experimental Study. *Merrill-Palmer Quarterly, 33*, 321-363.

Watkins, M., & Edwards, V. (2004). Assessing Early Literacy Skills with the Mountain Shadows Phonemic Awareness Scale (MS-PAS). *Journal of psycho-educational assessment,22,3-14.*

Weil, M. J., & Amundson, S. J. (1994). Relationship between visuomotor and handwriting skills of children in kindergarten. *American Journal of Occupational Therapy, 48*(1), 982-988.

Wilson, S. & Lonigan, C. (2010). Identifying Preschool Children at Risk of Later Reading Difficulties: Evaluation of Two Emergent Literacy Screening Tools. *Journal of Learning Disabilities, 43, 62-76*

Whitehurst, G. J., & Fischel, J. E. (1994). Practitioner review: Early developmental language delay: What if anything should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35*(4), 613-648.

Wright, R., Martland, J., & Stafford, A. (2006). *Early numeracy: Assessment for teaching and intervention*. (2nd ed.). London: Paul Chapman Publishing.