THE USE OF THE ARABIC CBM MAZE AMONG THREE LEVELS OF ACHIEVERS IN JORDAN

Bashir Abu-Hamour
Mutah University

This study examined the applicability of the Arabic version of the Curriculum Based Measurement Maze (CBM Maze) for Jordanian students. A sample of 150 students was recruited from two public primary schools in Jordan. The students were ranked into high, moderate, and low achievers in terms of their performance in the Arabic course. Then all of them were administered the Arabic CBM Maze probes. The students' scores in the Arabic CBM Maze were less than the previous American studies and norms. The results indicated that the Arabic CBM Maze is a reliable, valid, and cost effective measure. In addition, the Arabic CBM Maze is a good predictor of the Arabic language Grade Point Average. Moreover, it can be concluded that the Arabic CBM Maze may be used with confidence to differentiate the students' levels of reading achievement.

Reading skills deficits are a common characteristic of students referred for special education services (Daly, Chafouleas, & Skinner, 2004; Lentz, 1988; Winn, Skinner, Oliver, Hale, & Ziegler, 2006). According to the report released by the National Assessment for Educational Progress (NAEP) in reading, 43% of fourth graders cannot read at the basic literacy level (Daane, Campbell, Grigg, Goodman, & Oranje, 2005). Reading receives a great amount of attention because students require skills in reading comprehension to access information and concepts in various curriculum areas (Brown-Chidsey, Davis, & Maya, 2003). Thus, students who display poor reading skills are more likely to experience difficulties in other academic areas, such as history, geography, and economics (Espin & Deno, 1993). These reading deficits are likely contribute to unsuccessful outcomes for students, such as high dropout rates, grade retention, and overall poor achievement (Malmgren, Edgar, & Neel, 1998; Wagner, D'Amico, Marder, Newman, & Blackorby, 1992).

However, when a child's reading problems are recognized early, school failure can, to a large extent, be prevented or reduced (Raikes et al., 2006). Early intervention to prevent development of reading difficulties can be an effective way to ameliorate this problem (Torgesen et al., 1999), and screening and progress monitoring can identify students who require such intervention (Compton, Fuchs, Fuchs, & Bryant, 2006). In view of this fact, it is of critical importance to have a valid and reliable assessment instrument to be used in identifying students who are at-risk of reading failure.

A commonly used and well-researched method for assessing students' reading is the curriculum based measurement (CBM). The CBM is considered to be a type of authentic assessment practice that is designed to provide prevention and intervention services to students (Hoover & Mendez-Barletta, 2008).The CBM's validity and reliability are well established (National Center on Response to Intervention, 2010). The CBM is a set of standardized procedures that were initially designed to index the level and rate of student achievement within the basic skill areas of reading, mathematics, written expression, and spelling (Deno, 1985; Deno, 2003). Researchers indicate that the CBM can provide accurate information about a student's academic standing and progress, which can then be used for a variety of psycho-educational decisions that include: (a) identifying students for special services (Fore, Burke, & Martin, 2006; Marston, Mirkin, & Deno, 1984; Shinn, 1989); (b) formulating goals and objectives for Individualized Educational Plans (IEPs; Deno, Mirkin, & Wesson, 1984); (c) monitoring student progress and improving educational programs (Fuchs, Deno, & Mirkin, 1984); (d) transitioning students to less restrictive environments (Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994); (e) evaluating school programs (Germann & Tindal, 1985); and (f) predicting how well students will perform on

statewide competency tests of achievement (Crawford, Tindal, & Stieber, 2001; Fore, Boon, & Martin, 2007).

In the area of reading, two types of CBM measures have been used in research and practice: the CBM oral reading fluency (ORF) and the CBM Maze. On the CBM ORF measure, student performance is measured by requiring students to read aloud passages of meaningful text for one minute. The number of words read correctly is scored as the reading rate (Deno, 1985). Although assessment of ORF is the primary CBM of reading used in research and practice (Reschly, Busch, Betts, Deno, & Long, 2009), the CBM Maze is growing in popularity as an additional measure. On the typical CBM Maze tasks, students are presented with a passage of approximately 250 words in which every seventh word has been deleted and replaced with three options. The increased use of the CBM Maze is partly due to efficiency of administration and because teachers perceive it as more reflective of reading comprehension than the ORF (Fuchs & Fuchs, 1992; Fuchs, Fuchs, & Maxwell, 1988). Recently, the CBM Maze has been receiving more attention due to the fact that it can be administered to a group of students at one time, whereas the CBM ORF is individually administered. Because the CBM Maze is group administered, an entire classroom or even an entire grade level can be assessed in less than five minutes. In addition to being potentially more efficient, the CBM Maze task might be more appropriate than the CBM ORF for use in screening for students in the intermediate (e.g., fourth and fifth) grades. After third grade, the primary emphasis of reading instruction switches from fluency to comprehension, and this switch may be reflected in the choice of universal screening measures.

The CBM Maze
The CBM Maze is a widely used assessment system for the universal screening of academic skills. Universal screening programs assess all students in a population (e.g., classroom, school, or district) with the intent of identifying those who are not making sufficient progress and addressing their academic needs with research-based interventions. The CBM Maze can be useful as a screening tool only if it differentiates readers by ability. The CBM must provide a reliable indicator of a student's overall proficiency in the academic skill of concern (e.g., reading).Because considerations of reliability and validity, time involved in assessment, and sensitivity to differences also are key considerations in selecting universal screening measures, many schools find that the CBM Maze is a useful screening tool.

In regard to the psychometric properties, the CBM Maze has been shown to provide a valid and reliable measurement of reading skills in elementary-, middle-, and high-school students (Brown-Chidsey et al., 2003; Espin & Foegen, 1996; Fuchs & Fuchs, 1992; Miura-Wayman, Wallace, Ives-Wiley, Ticha, & Espin, 2007; Shinn, Deno, & Espin, 2000). Results from previous research have indicated that the Maze has adequate technical characteristics, is sensitive to improvement of student performance over a school year, and can reveal inter-individual differences in growth rates (Shin et al., 2000). Moreover, several studies support the alternate form, reliability, sensitivity to growth, and predictive validity of the CBM Maze (e.g., Espin, Wallace, Lembke, Campbell, & Long, 2010; Graney, Martínez, Missall, & Aricak, 2010; Shin et al., 2000) has been established. In addition, CBM Maze has been found to correlate with state accountability tests (Fore et al., 2007).

With the emphasis on accountability, a growing focus is to use the CBM to predict student performance on state competency tests of achievement (Tindal & Marston, 1990). Tindal et. al., (2003) indicated that predicting student performance on statewide competency tests of achievement is critical. More efficient measures that can provide similar information can be extremely valuable for teachers. Measures that give teachers snapshots of students' conceptual understanding of academic concepts at their grade level can fill the need for formative progress monitoring. In addition, justification for predicting achievement scores can be found in the school accountability movement that has put a premium on educators' providing evidence of student learning (Ysseldyke, Thurlow, & Shriner, 1992). For the purpose of this study, it is expected that predicting students reading in Arabic will present many obstacles due to the complex nature of the Arabic orthography.

The Challenges of the Arabic Language
Several graphical features of the Arabic language create certain difficulties in learning and teaching reading skills. First, Arabic is an alphabet language with 28 letters, written in a joined fashion from right to left (Abu-Rabia & Siegel, 2002). All letters are consonants except three long vowels. Another three short vowels (diacritics) do exist in the form of separate diacriticals, not as independent graphemes. When any of these diacritics appear on certain letters, it gives the letter a completely different sound; for example, the letter *k* could have any one of the sounds *ka*, *ki*, or *ku*. If the same letter *k* comes in a word

where it does not need a vowel, its sound will be *ek*. Therefore, when these diacritics or short vowels appear in the script Arabic shows a high degree of regularity and the students can read by predicting the sound of the letters. However, in most modern and printed Arabic text (grade four and above) vowel signs are not given or given partially, therefore reading relies more on the context rather than spelling and Arabic script becomes more irregular (Abu Rabia, 2002; Abu Rabia & Siegel, 2002). Second, the Arabic script is written in a cursive fashion while each individual letter has multiple forms or shapes according to its position within the word. Many letters, furthermore, have similar graphemes but their phonemes are completely different. The Arabic alphabet consists of letters with almost twenty letters having graphic similarity with at least one or two other letters (Brenznitz, 2004). Third, a greater influence of orthographic processing over-and-above phonological processing could be related to diglossia (the existence of a formal literary form of a language along with a colloquial form used by most speakers) in Arabic. Saiegh-Haddad (2007) has argued that differences between the spoken form of Arabic experienced by the preschool child (e.g., a local dialect) and the standard form of Arabic used in education and writing disrupts the construction of phonological representations of Arabic. Fourth, the glottal stop in Arabic, referred to as the Hamza, although a fully functioning consonant, is treated as a diacritical mark and has many different ways of writing depending on its position in the word resulting in various complex spelling and reading conventions (Elbeheri, Everatt, Mahfoudhi, Abu Al Diyar, & Taibah, 2011). Finally, the Shaddah, one of the <u>diacritics</u> used with the <u>Arabic alphabet</u>, is marking a long consonant. Shaddah is not a vowel. It indicates a place where the written language is showing only one consonant, but you are expected to pronounce two consonants. Normally, this means that you have to hold (sustain) the sound of that letter for twice as long as you normally would.

With all of the challenges of teaching and learning Arabic, it is a necessity to explore valid and reliable measures that can be used for predicting reading skills and identifying students with reading difficulties in the Arab world. This study is intended to investigate the applicability of the CBM Maze procedure in the Arabic language.

Significance of the Study
The main aim of most tests is to determine the academic levels of the students, particularly exceptional students who are far behind or far ahead of classmates. In Arab countries very limited research exists that addresses effective assessment practices for students who are severely deficient in reading or superior in reading (Al-Mannai & Everatt, 2005). The difficulty and complexity of the orthography of the Arabic language may explain the need to validate a screening and progress monitoring tool such as the CBM Maze test in Arabic to predict reading skills in the early stages of schooling. The educational systems in the Arab countries lack valid and reliable assessment tools that can be used to identify students who are at risk of developing reading difficulties (Al-Mannai & Everatt, 2005; Elbeheri et al., 2011). For example, researchers in Jordan have stated in numerous reports and articles that the Jordanian educational system is in need of valid assessment tools to identify students with reading disability and provide them with appropriate interventions (Al-Khateeb, 2008; Al-Natour, 2008).

Students with reading difficulties need a classroom-based measure of reading that is sensitive, efficient, and otherwise acceptable to teachers. The literature base on the CBM Maze measure is well established. However, there is a need for an examination on the use of the CBM Maze for students who speak languages other than English. Specifically, the CBM tools need to be validated in the Arabic language. Developing a formal assessment tool that can be used to find students with reading difficulties then follow their progress is a critical need in Jordan as well as other Arab countries. Students who have special needs in the Arab world are usually expelled or drop out from public schools because early adequate service and assessment are not provided to help them succeed. There is a need for a screening and progress monitoring instrument for the purpose of identifying at-risk children at time of school entry and providing identified children with systematic interventions (Al-Khateeb, 2007, 2008; Al-Natour, 2008; McBride, 2007).When a child's problems are recognized early, school failure can to a large extent be prevented or reduced (Raikes et al., 2006). To the author's knowledge, no studies have been conducted to investigate the applicability, reliability, and validity of the CBM Maze measure with Arabic speaking children.

Purposes of the Study
The purposes of this study were to explore the CBM Maze applicability, reliability, and validity with three levels of Jordanian students who speak Arabic. This study addressed the following three major questions:

Study Question 1: To what extent will the Arabic CBM Maze be a reliable measure of reading ability among three levels of achievers?
Study Question 2: What is the relationship between the Arabic CBM Maze and the Arabic Language Grade Point Average among three levels of achievers?
Study Question 3: To what extent do high achievers, moderate achievers, and low achievers differ in their Arabic CBM Maze scores?

Method
Participants
The sample of the study consisted of 150 students from fourth grade. These students were divided into three sub-samples. These samples were classified into high achievers, moderate achievers, and low achievers in their Arabic course with 50 participants for each one. The Arabic teachers and the researcher of this study ranked the participants into high, moderate and low achievers. Both the teachers' judgment and the students Arabic GPA of the first semester of 2011/2012 were used to group the students. The Arabic GPA cutoff points for classifying students were 69 and below for low achievers, 70 to 85 for moderate achievers, and 86 and above for high achievers. Then, all participants were chosen randomly and consent forms were sent to parents seeking their agreement of participation. Parents who agreed to let their children participate in the study were requested to complete a short questionnaire that addressed the inclusion criteria of this study. The participants were selected from a larger set of students (446) who were assessed to meet the requirements for inclusion in the study: intelligence within the average range, native speakers of Arabic, no noted emotional or behavioral disorder, no noted attention disorders, and no sensory impairments. The mean age of the participants was nine years and eight months with a range of 115-119 months. Of the total sample, 75 were male and 75 female. These students enrolled in the second semester of 2011/2012. All participants were administered the Arabic CBM Maze probes. Participants were recruited from two public primary schools in a college town in the southern region of Jordan.

The data collection was completed by the researcher and two trained teachers residing in the southern region of Jordan. These teachers have a degree in special education and childhood education. The measures of the study, the Arabic CBM Maze instructions translation, and reliability are described in the following sections.

The Arabic language GPA. The Arabic GPA reflects a student's ability on three basic Arabic skills: reading (word reading and reading comprehension), writing, and spelling in the accredited Arabic curriculum in Jordan. The Arabic GPA is a numeric average of all grades achieved in classes at a given school semester. The purpose of GPA is to provide a barometer as to overall performance of a student in his or her classes, as well as create a system that allows for comparisons between students, and a class ranking system. In the Jordanian educational system, students are ordered and assigned a numerical rank against their peers based on their GPA, starting with number 100 for the student with the highest GPA and 0 for students with the lowest GPA. The rubric for the Arabic GPA is excellent (90-100), very good (80-89), good (70-79), satisfactory (60-69), minimal pass (50-59), and failure (< 50). In this research, the mean Arabic GPA of the high achievers was 89.9 with a range of 86 to 98 and standard deviation of 4.50. For moderate achievers, the mean was 78.7 with a range of 70 to 85 and standard deviation of 4.33. The mean was 61.9 with a range of 50 to 69 and standard deviation of 6.52 for low achievers.

Translating the CBM Maze instructions into Arabic. The researcher used appropriate translation procedures (Brislin, 1986) prior to administer the Arabic CBM Maze to a sample of Jordanian students. First, two native speakers of Arabic, who were also fluent in English, independently translated the CBM Maze instructions into Arabic. Second, a back translation of the Arabic version into English by a bilingual resident of the United States who is fluent in both English and Arabic languages was conducted. Third, all translators reached a reconciliation of the forward-backward translations. Finally, a pretest was conducted with a convenience sample of 15 fourth grade students to assess ease of comprehension, possible ambiguities, and alternative administration wording.

The CBM Maze probes. The content of the probes were selected from several literature-based reading series used in the educational system in Jordan as supplementary materials to the accredited curriculum of fourth grade. Using a table of random numbers to select page numbers within books, potential passages were examined and excluded if they contained excessive dialogues, poetry, plays or many unusual or foreign names. Moreover, all probes were partially vowelized just like the accredited curriculum that students are exposed to. Then a pool of probes was selected by the author and the Arabic curriculum specialist who works in the curriculum department in the Ministry of Education in Jordan.

Each probe includes approximately 300 words. Researchers have tried to reduce variability in individual students' data due to passage difficulty by using readability formulas to measure text difficulty (Griffiths, VanDerHeyden, Skokut, & Lilles, 2009). In this study, the researcher used the Spache formula to reduce the variance of the scores on the Arabic CBM Maze (Good & Kaminski, 2002). This formula considers difficulty of vocabulary and sentence length. In addition, all probes were given to three university's instructors and three teachers in the field to judge the difficulty of grammar and word order. Their suggestions were taken into consideration to make the final version of the probes (see ''Appendix'' for the Arabic CBM Maze sample). The procedure was identical to the one used with English CBM Maze probes.

To save set-up time and obtain a more accurate score, three equivalent probes of the Arabic CBM Maze materials were administered to the group of students in one testing session. The median score of these three probes were used to provide the valid data point on the student's performance. The researchers counted the total number of responses attempted in three minutes and the total number of errors then subtracts the total number of errors from the total number attempted. Their performance is then based on the Words Correctly Restored (WCR) score (Hosp et al., 2007). To establish test-retest reliability, other three equivalent probes of the Arabic CBM Maze materials were administered to the same group of students in the second day and the median scores were used to represent their performance.

Procedural and inter-rater reliabilities. To ensure consistency of testing administration across the CBM probes, the researchers read from scripts and used timers. The fidelity of testing administration was tested by using a detailed checklist to ensure each test was administered as it was intended and described in the manuals of the CBM testing (Hosp et al., 2007). Procedural reliability was obtained during 100% of testing sessions with an average reliability of 100 percent. The teachers scored each CBM Maze probe and entered the data into an excel sheet. The researcher checked randomly 30% of the scoring sheets. The average inter-rater reliability of scoring fidelity data was 99% (range 98%-100%). In terms of data entry reliability, all of the excel data (100%) were checked against the paper scores and all discrepancies were resolved by examining the original protocols.

Results

Preliminary Data Analysis

First of all, to improve the shape of the distributions, the responses of outliers whose scores were ±2 standard deviation or more from the group mean were replaced by a value equal to the next highest non-outlier-score plus 1 unit of measurement (Tabachnick & Fidell, 2001). Table 1 presents the descriptive analysis included calculating the means, standard deviations, and percentile ranks among high achievers, moderate achievers, low achievers and all achievers of the Arabic CBM Maze performance in WCR. This descriptive information was helpful in understanding the data and making initial inferences on the differences among all groups. Descriptive statistics also allowed providing visual graphs that facilitated more convenient presentation of the data. Graph 1 displays the average performance of the Arabic CBM Maze of the three students' levels. In general, the preliminary results indicate differences among all groups. A closer inspection of the data analysis that addressed study's questions is followed.

Figure 1. Graphic display of the mean performance on the CBM Maze measure reported in Words Correctly Restored in three minutes.

Table 1. Descriptive Information of the CBM Maze Performance in WCR for All Groups

Group	Number of Students	Range	M	SD	Percentile	
High Achievers	50	23-32	27.4	2.06	90%	30
					75%	29
					50%	27.5
					25%	26
					10%	24.1
Moderate Achievers	50	13-24	18.48	2.74	90%	22
					75%	21
					50%	18
					25%	16.75
					10%	14.10
Low Achievers	50	4-13	8.84	2.34	90%	11.90
					75%	10
					50%	9
					25%	7
					10%	5.10
All Achievers (Students)	150	4-32	18.24	7.96	90%	29
					75%	22
					50%	15
					25%	10
					10%	6.10

Note. WCR = Words Correctly Restored; M = Mean; SD = Standard Deviation.

The Arabic CBM Maze Test-Retest Reliability
The median baselines scores of the students' performance on the two consecutive days were correlated to establish test-retest reliability. The correlations coefficients were .84, .85, .83, and .89 for low achievers, moderate achievers, high achievers, and all achievers respectively. These resulting coefficients are large enough to demonstrate that the Arabic CBM Maze has acceptable test-retest reliability. In addition, The Standard Error of Measurements (SEMs), reported in Table 2, can be used to estimate the confidence interval that surround a particular CBM Maze score. The SEM is based on the formula SEM= SD square root 1-*r*; (SD-Standard Deviation and r- reliability) and establishes a zone within which an individual's true score probably lies. The smaller the SEM, the more confidence one can have in the test's results. Arabic CBM Maze has small SEMs (range from .84 to 1); examiners can use it with confidence.

Table 2. Test-Retest Reliability and SEMs for the Arabic CBM Maze

	First Testing		Second Testing			
Grade Level of Sample	M	SD	M	SD	*r*	SEMs
High (n = 50)	27.40	2.06	28.26	1.92	.83	.84
Moderate (n = 50)	18.48	2.74	18.72	3.09	.85	1
Low (n = 50)	8.84	2.34	9.54	2.30	.84	.93
All Achievers (n = 150)	18.24	7.96	18.84	8.05	.89	.87

Note. M = Mean; SD = Standard Deviation; r = Correlation Coefficient; SEMs= Standard Errors of Measurement.

The Relationship between the Arabic CBM Maze and the Arabic GPA
The Arabic CBM Maze scores were correlated with the Arabic GPA for all participants and in all levels. All of the coefficients were statistically significant at the *p* < .05 level. They range in magnitude from moderate (for moderate achievers) to large (for all achievers). The correlations coefficients were .40, .32, .35, and .81 for low achievers, moderate achievers, high achievers, and all achievers respectively.

The Average Arabic CBM Maze Differences among Three Levels of Achievers
To explore differences among the three levels of achievers, one-way independent Analysis of Variance (ANOVA) was performed. All assumptions of performing ANOVA were examined. No violations of normality and homogeneity of variance were detected. The variances were equal for all three groups, $F(2, 147) = 2.73$, $p > .05$. There were significant differences among the three groups of achievers in terms of their CBM Maze scores, $F(2, 147) = 749.27$, $p<.001$, $\omega = .90$. In addition, there was a significant linear trend, $F(1, 147) = 1497.80$, $p<.001$, $\omega = .90$, indicating that as the level of achievement increased, the CBM Maze increased proportionately.

Discussion
The primary purpose of this study was to broaden the knowledge base regarding the applicability of the CBM Maze assessments in Arabic language with Jordanian students. Very rigorous steps were performed to assure the accurate translation of the CBM Maze instructions. In addition, procedures were taken to assure that all Arabic CBM Maze probes were equivalent in difficulty. It can be concluded that compared to the CBM Maze from AIMS web (2008), Jordanian students restored fewer number of correct words. This can be attributed to the fact that speed reading within three minutes is a new practice for them. It may be the case that the culture-related content of the literary works had an effect on the students' reading scores. In addition, some characteristics of the Arabic system may result in great difficulty for children reading Arabic. Most of these factors or characteristics are related to the orthographic features of Arabic language (for review see Abu Rabia, 2002; Abu Rabia & Siegel, 2002; Breznitz, 2004). The study's questions will be discussed in the following sections.

The Arabic CBM Maze Reliability
The Arabic CBM Maze reliability was investigated by procedural, inter-rater, and test-retest reliabilities. The resulting coefficients were very high for procedural and inter-rater reliabilities. Although test-retest reliability is high enough as well, a sizable proportion of the variance in scores was attributable to overall mean differences in performance across probes, most likely reflecting differences in difficulty across probes. These findings mirror reliability coefficients for the CBM maze presented in previous research (Shin et al., 2000). In addition, very small SEMs were detected in this study which leads to the conclusion that the Arabic CBM Maze scores are consistent across a short period of time and across different examiners.

The Relationship between the Arabic CBM Maze and the Arabic GPA
Although the Arabic GPA cannot be considered a standardized assessment due to the certain degree of subjective judgment that teachers should made about students reading ability, the use of it was imperative since no standardized assessment was existed in Arabic to be used for the purpose of the study. Examinations of criterion validity between the Arabic GPA and the Arabic CBM Maze yielded interesting results because relationships were significant and high when merge all level of achievers. The significant correlation between the Arabic CBM Maze and the Arabic GPA aligns with previous research documentation of correlations existing between this measure and standardized assessments (Fewster & Macmillan, 2002; Fore et al., 2007; Shin et al., 2000). Predicting student performance on Arabic competency tests of achievement is critical. More efficient measures such as the Arabic CBM Maze that provide similar information can be an extremely valuable tool for teachers. The results of this study indicated that the Arabic CBM Maze can be used to inform language outcome that including reading comprehension for students in fourth grade. These results support other researchers' assertions that an important relationship exists between academic language proficiency and reading skill acquisition (e.g., Thomas & Collier, 2002).

The Use of Arabic CBM Maze as Universal Screening Tool
In this research, significant differences were found among the three groups of achievers in terms of their Arabic CBM Maze scores. In addition, there was a significant linear trend indicating that as the level of achievement increased, the Arabic CBM Maze increased proportionately. This study suggests that the existing Arabic CBM Maze measure may be adequate when group administration is necessary or desirable for universal screening so long as multiple probes are collected per occasion to rank the students and identify students who will struggle in reading.

In conclusion, the CBM Maze has been shown to be a valid and reliable measurement of reading skills in elementary-, middle-, and high-school students (Brown-Chidsey et al., 2003; Espin & Foegen, 1996; Fuchs & Fuchs, 1992; Miura-Wayman et al., 2007; Shinn et al., 2000). In this line, the results of this

research have indicated that the Arabic CBM Maze has adequate technical characteristics. Arabic CBM Maze has been found to correlate with fourth students' Arabic GPA and can be used as universal screening tool to identify exceptional students who are far behind or far ahead of classmates in reading comprehension.

Limitations, Future Research, and Implications
This study has several limitations that should be considered. First, data were only collected on fourth-grade students; consequently, the generalizability of findings to other grades is unknown. Second, the sample size was relatively small and came from public school district. Third, the study examined concurrent relationships at one point in time. Future studies would need to be conducted that examined related validity factors such as sensitivity to progress and predictive validity. Additional development and field testing of the Arabic CBM Maze probes is recommended prior to more widespread use of the CBM Maze for absolute decisions (e.g., comparing specific scores to cutoffs or progress monitoring for individual students). Future research should further compare the instruments with other student populations to evaluate superiority with regard to efficiency in administration and scoring and to predicting high-stakes outcomes.

Teachers are challenged to meet the wide range of needs of an increasingly diverse student body while at the same time ensuring that all are progressing toward high academic standards. It is thus critical that they have the means to identify students who are not making sufficient progress toward those standards and to make instructional decisions based on technically sound data. This study indicates that the CBM Maze, which has been demonstrated to be reliable and valid measures for many students who speak English, also shows promise in measuring reading comprehension in Arabic. Teachers in the Arab world should consider other valid and reliable assessment tools such as CBM Maze for use in both general and special education systems. Particularly important aspects of CBM Maze for use in Arabic speaking countries are the ease of administration, the low cost, and the reliability of the measure.

Appendix. Sample of the Arabic CBM Maze Probe

وادي الجواهرِ

في إحدى الرّحلاتِ البحريّةِ للسّندبادِ هبّتْ على السّفينةِ الّتي كانتْ تحملُه عاصفةٌ شديدةٌ، ولمْ يتمكّنْ (**صاحبُ، قائدُ، صانعُ**) السّفينةِ والبحّارةُ منَ السّيطرةِ عليها، فقذفتْهم (**الرّياحُ، الغيومُ، الأمواجُ**) العاليةُ إلى ساحلِ جزيرةٍ بعيدةٍ، فنزلوا فيها، وكانتْ (**سماؤُها، أرضُها، ماؤُها**) مغطّاةً بالأشجارِ. وسارَ السّندبادُ بينَ (**المنازلِ، الأشجارِ، الصّخورِ**) بعيدًا عن الآخرينَ، فرأى شجرةً غريبةً، فأكلَ منْ (**أغصانِها، أوراقِها، ثمارِها**) دونَ أنْ يعلمَ أنَّ هذهِ الثّمارَ تسبّبُ النّومَ، فراحَ في (**نومٍ، نشاطٍ، حزنٍ**) عميقٍ. أفاقَ السّندبادُ بعدَ نومٍ طويلٍ، فركضَ نحوَ (**الجبالِ، الرّمالِ، الشّاطئِ**) مذعورًا، فلمْ يجدْ السّفينةَ وبقيَ وحيدًا في هذهِ (**الجزيرةِ، القريةِ، المدينةِ**) الموحشةِ، ولمّا حلَّ الظّلامُ زادَ خوفُهُ، ولكنّهُ لمْ ييأسْ، وواصلَ (**الأكلَ، النّومَ، السّيرَ**)، ولمّا أدركَهُ التّعبُ نامَ.

References
Abu-Rabia, S. (2002). Reading in a root-based morphology language: The case of Arabic. *Journal of Research in Reading, 25*, 320-330.
Abu-Rabia, S., & Siegel, L. S. (2002). Reading, syntactic, orthographic, and working memory skills of bilingual Arabic-English speaking Canadian children. *Journal of Psycholinguistic Research, 31*, 661–678.
AIMSweb. (2008). *Curriculum based measurement norms* [Date file]. Available at http://www.aimsweb.com
Al-Khateeb, J. (2007). *Final report: Special needs education 3.3c program development–mild intellectual disabilities*. Prepared for the Ministry of Education. The Hashemite Kingdom of Jordan under the ERFKE 1 project.
Al-Khateeb, J. (2008). *Final report 3.4b program evaluation: Resource rooms*. Prepared for the Ministry of Education. The Hashemite Kingdom of Jordan, ERFKE 1 Project.
Al-Mannai, H. A., & Everatt, J. (2005). Phonological processing skills as predictors of literacy among Arabic speaking Bahraini school children. *Dyslexia, 11*, 269–291.
Al-Natour, M. (2008). *Final report: Special needs education 3.3e program development–Learning Disabilities*. Prepared for the Ministry of Education. The Hashemite Kingdom of Jordan under the ERFKE 1 project.
Brenznitz, Z. (2004). Introduction on regular and impaired reading in sematic languages. *Reading and Writing: an Interdisciplinary Journal, 17*, 645-649.
Brislin, G. J. (1986). The wording and translation of research instruments. In W. L. Loner & J.W. Berry (Eds.), *Field Methods in Cross-Cultural Research* (pp. 137–164). Newbury Park, CA: Sage.

Brown-Chidsey, R., Davis, L., & Maya, C. (2003). Sources of variance in curriculum-based measures of silent reading. *Psychology in the Schools, 40*, 363-377.

Compton, D. L., Fuchs, L. S., Fuchs, D., & Bryant, J. D. (2006) Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology, 98*, 394-409.

Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7*(4), 303–323.

Daly, E. J. III, Chafouleas, S., &Skinner, C. H. (2004). *Interventions for reading problems: Designing and evaluating effective strategies*. New York: Guilford Press.

Daane, M.C., Campbell, J.R., Grigg, W.S., Goodman, M.J., & Oranje, A. (2005). *Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading* (NCES 2006-469). U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics. Washington, DC: Government Printing Office.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37*, 184-192.

Deno, S., Mirkin, P., & Wesson, C. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children, 16*(2), 99–104.

Elbeheri, G., Everatt, J., Mahfoudhi, A., Abu Al-Diyar, M., & Taibah, N. (2011). Orthographic processing and reading comprehension among Arabic speaking mainstream and LD children. *Dyslexia, 17*, 123-142.

Espin, C. A., & Deno, S. L. (1993). Performance in reading from content area text as an indicator of achievement. *Remedial and Special Education, 14*, 47 – 59.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures of second students' performance on content area tasks. *Exceptional Children, 62*, 497-515.

Espin, C., Wallace, T., Lembke, E., Campbell, H., & Long, J. D. (2010). Creating a progress-monitoring system in reading for middle-school students: Tracking progress toward meeting high-stakes standards. *Learning Disabilities Research & Practice, 25*, 60-75.

Fewster, S., & Macmillan, P. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23*(3), 149–156.

Fore, C., Boon, R., & Martin, C. (2007). Concurrent and predictive criterion-related validity of curriculum-based measurement for students with emotional and behavioral disorders. *International Journal of Special Education, 22*(2), 24–31.

Fore, C., Burke, M., & Martin, C. (2006). Curriculum-based measurement and problem-solving model: An emerging alternative to traditional assessment for African American children and youth. *Journal of Negro Education, 75*, 16–24.

Fuchs, L. S., Deno, S., & Mirkin, P. (1984). Effects of frequent curriculum based measurement and evaluation on pedagogy, student achievement and student awareness of learning. *American Educational Research Journal, 21*, 449–460.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*, 45-58.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*, 20-28.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60*(6), 518–537.

Germann, G., & Tindal, G. (1985). An application of curriculum based assessment: The use of direct and repeated measurement. *Exceptional Children, 52*(3), 244–265.

Graney, S. B., Martínez, R. S., Missall, K. N., & Aricak, O. T. (2010). Universal screening of reading in late elementary school: R-CBM versus CBM Maze. *Remedial and Special Education, 31*, 368-377.

Griffiths, A. J., VanDerHeyden, A. M., Skokut, M., & Liles, E. (2009). Progress monitoring oral reading fluency within the context of RTI. *School Psychology Quarterly, 24*, 13-23.

Good, R. H. & Kaminski, R. A. (2002). *DIBELS Oral Reading Fluency Passages for First through Third Grades* (Technical Report No. 10). Eugene, OR: University of Oregon.

Hoover, J. J., & Mendez-Barletta, L. M. (2008). Considerations when assessing ELLs for special education. In J. K. Klingner, J. J. Hoover, & L. Baca (Eds.). *Why do English language learners struggle with reading* (93-108). Thousand Oaks, CA: Corwin Press.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum based measurement.* New York: Guilford.

Lentz, F. E. (1988). Effective reading interventions in the regular classroom. In J. L. Graden, J. E. Zins, & M. J. Curtis (Eds.), *Alternative educational delivery systems: Enhancing instructional options for all students* (pp. 351 – 370). Washington, DC: National Association of School Psychologists.

Malmgren, K., Edgar, E., & Neel, R. S. (1998). Postschool status of youths with behavioral disorders. *Behavioral Disorders, 23*, 257 – 263.

Marston, D. B., Mirkin, P. K., & Deno, S. L. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *Journal of Special Education, 18(*2), 109–117.

McBride, S. (2007). *Review of special education –Grades 1 to 12*. Summary Report Submitted to the Ministry of Education, Amman, Jordan.

Miura-Wayman, M., Wallace, T., Ives-Wiley, H., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41*, 85 – 120.

National Center on Response to Intervention. (2010). *Essential components of RTI: A closer look at response to intervention*. Retrieved from http://www.cldinternational .org/Articles/rtiessentialcomponents.pdf

Raikes, H. H., Torquati, J. C., Hegland, S., Raikes, H. A., Scott, J., Messner, L. (2006). Studying the culture of quality of early education and care. In M. Zaslow & I. Martinez-Beck (Eds.). *Critical issues in early childhood professional development* (pp.111–136). Baltimore. MD: Paul H. Brookes.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-Based Measurement Oral Reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*, 427-469.

Saiegh-Haddad, E. (2007). Linguistic constraints on children's ability to isolate phonemes in Arabic. *Applied Psycholinguistics, 28,* 607-625.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children*. New York, NY: Guilford Press.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *The Journal of Special Education, 34*, 164 – 172.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston, MA: Allyn & Bacon.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Santa Cruz, CA: Center for Research on Education, Diversity & Excellence.

Tindal, G., & Marston, D. (1990). *Classroom-based Assessment: Evaluating Instruction Outcomes*. Columbus, OH: Merrill Publishing Company.

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children, 69*(4), 481-494.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., & Garvin, C. (1999). Preventing reading failure in your children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91*, 579-593.

Ysseldyke, J., Thurlow, M., & Shriner, J. (1992). Outcomes are for special educators too. *Teaching Exceptional Children, 25*(1), 36-50.

Wagner, M., D'Amico, R., Marder, C., Newman, L., & Blackorby, J. (1992). *What happens next? Trends in postschool outcomes of youth with disabilities*. The second comprehensive report from the National Longitudinal Transition Study of Special Education Students. Menlo Park, CA: SRI International.

Winn, B. D., Skinner, C. H., Oliver, R., Hale, A. D., & Ziegler, M. (2006). The effects of listening-while-reading and repeated reading on the reading fluency of adult learners. *Journal of Adolescent and Adult Literacy, 50*, 196 – 205.