An Examination of the Diagnostic Utility of Three Dyscalculia Identification Models

Elias Clinton
Black Hills State University

Abstract
Dyscalculia is a term used to describe a learning disability associated with difficulty in using, understanding, and manipulating numbers. Dyscalculia is an evident and consistent underachievement in mathematics that cannot be explained by health, sensory, intelligence, environmental factors, or as a result of poor instruction. Currently there is a debate amongst scholars regarding the most accurate model for diagnosing dyscalculia. Given the ongoing debate regarding dyscalculia diagnosis, this review of current scholarly literature was conducted to evaluate the diagnostic utility of three models of dyscalculia identification: Aptitude-Achievement Discrepancy (AAD), Response to Intervention (RTI) Low Achievement (LA). A summary of each relevant study is provided, the diagnostic utility of each identification model is discussed, and implications for future research are made. Future implications include a need for research that further explores the underlying cognitive processing deficits associated with dyscalculia, the need for extensive longitudinal studies that examine the persistence of dyscalculia throughout all grade levels, and a recommendation of a hybrid model of dyscalculia classification.

Keywords: dyscalculia, mathematical disorders, learning disabilities, acalculia, identification, diagnosis, classification, discrepancy, achievement, RTI.

1. Introduction
The Individuals with Disabilities Education Act (IDEA) of 2004 defines a learning disability (LD) as a “disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations (IDEA, 2004). Students with LD constitute the largest portion of special education (50 percent) throughout the world, with an estimated 4 - 7 percent of students in the United States being served under the category (Büttner & Hasselhorn, 2011) [1].

Despite a federal definition of LD, decades of extensive research, and a number of distinguished scholars who have devoted their lives to studying the construct of LD, controversial debate continues to grow surrounding the identification of LD. This debate is apparent across all LD subgroups: math LD (dyscalculia), reading LD (dyslexia) and written expression LD (dysgraphia); however the area of math LD is particularly divided resulting in a large gap in the literature. The lack of empirical data regarding math LD may be due to the lack of an agreed upon definition, complex conceptual issues, and the intricacy of the numerous cognitive processes associated with mathematical ability (Landerl, Bevan, & Butterworth, 2004; Mazzocco & Myers, 2003) [9, 12].

“Mathematical disorders” and “math learning disabilities” are broad terms that include a wide variety of student characteristics and a multitude of difficulties associated with mathematics (e.g., computations, concepts, reasoning). In order to simplify the multifaceted construct of mathematical disorders, this review will use the all-encompassing term “dyscalculia.” Dyscalculia is a term used to describe a learning disability associated with difficulty in using, understanding, and manipulating numbers (Cohen, Soskic, Iuculano, Kanai, & Walsh, 2010) [2]. A common misunderstanding is that an individual experiencing difficulty with the complex nature of math (e.g., trigonometry, calculus, algebra) is an indication of dyscalculia, yet this is not the case (Cohen et al., 2010) [2]. Dyscalculia is an evident, yet unexpected, consistent underachievement in mathematics that cannot be
Dyscalculia is included under the category of Specific Learning Disability (SLD: IDEA, 2004), yet students are rarely identified and evaluated for special education services based solely on mathematical difficulties (Büttner & Hasselhorn, 2011) \[1\]. Approximately 85 percent of students identified under the category of SLD are served based on poor reading achievement (Büttner & Hasselhorn, 2011) \[1\]. Not surprisingly, the amount of existing research regarding math LDs is far less than that of reading LDs (Hanich, Jordan, Kaplan, & Dick, 2001; Mazzocco & Myers, 2003) \[8, 12\]. In fact, the research gap is so large that no core deficit has been pinpointed and no formal definition of dyscalculia has been established (Mazzocco & Myers, 2003) \[12\].

Due to the disproportionality of identified reading disorders to mathematical disorders, the prevalence and magnitude of mathematical disorders is not widely known. An estimated 6 percent of school-aged children in the United States are reported as having significant mathematics deficits (Mazzocco & Myers, 2003) \[12\]. Therefore, based on the widespread nature of mathematical disorders there is an imperative need for an evidence-based method of classifying dyscalculia. People with dyscalculia suffer from lifelong arithmetic and numerical difficulties that can result in detrimental effects on their vocations and quality of life. In light of the potentially debilitating effects of dyscalculia, it is imperative to develop a valid system for identifying and serving those students with the disorder.

Some commonly used models of dyscalculia identification include: aptitude-achievement discrepancy (AAD), response to intervention (RTI), and low achievement (LA). Using the AAD model, dyscalculia is diagnosed when a significant difference exists between a child's aptitude (typically measured using an intelligence test) and actual level of math academic achievement (as measured by an achievement test). RTI is a process that emphasizes measuring a struggling student’s response to individualized, intensive math instruction. Teachers provide scientific, research-based interventions, monitor and measure progress, and then use those measures to help inform classification for dyscalculia. LA is a model in which students who demonstrate math achievement below a certain cutoff score on standardized measures (e.g., 10th or 25th percentile) are identified as having dyscalculia. The validity and reliability of each of the aforementioned identification models has been both criticized and praised by scholars in the fields of psychology, education, and public health. Currently there is no consensus as to which of the identification models is considered most effective. Each of the models has been used to determine special education services for students, yet the efficacy of each is highly disputed. Thus, the purpose of this review is to investigate the diagnostic utility of the three models most commonly used to identify dyscalculia.

Methods

Only articles from peer-reviewed journals were chosen for this review. Studies were located using a computer search of scholarly (peer-reviewed) literature from various online databases including: ERIC, EBSCOhost, Medline, PsycInfo, and Google Scholar. The terms used in the search included: dyscalculia, mathematical disorders, learning disabilities, acalculia, identification, diagnosis, classification, discrepancy, low achievement, and RTI.

Data Collection

Studies for the review were chosen based on the following criteria: the publication date was between the years 2000 and 2015, the study involved a quantitative measurement of one of the three identification models (AAD, RTI, LA), an adequate description of data collection and measurement procedures were included, the participants were students in K – 12 classrooms, and the publication was from a peer-reviewed journal. The references of each selected study were then examined for additional relevant studies.

Characteristics of the Data Set

From the initial pool of studies located using the specified search methods, 72 articles were identified as being relevant to this review. It should be noted that the majority of studies located through the initial search focused on the broad topic of learning disabilities or reading disabilities (e.g., dyslexia). Ten of the 72 studies were empirically-based examinations of both dyscalculia and of one of the three targeted identification models. Each publication was examined and placed into a category based on the type of LD identification model pinpointed within the study. Of the ten empirical studies, five examined the LA model, two compared the LA model to the AAD model, one study examined solely the AAD model, one study compared LA to RTI, and one study compared all three of the models (RTI, LA, and AAD).

Results

Aptitude-Achievement Discrepancy (AAD) Model

Mazzocco and Myers (2003) \[12\] conducted a longitudinal study that compared the LA and AAD models. The participants in the longitudinal study included 209 students that the researchers assessed and monitored from Kindergarten to third grade. Participants qualified for dyscalculia using the LA model if their standard scores on the Test of Early Mathematics Ability – 2nd Edition (TEMA-2) fell at or below the 10th percentile (compared to other participants in the study). Students qualified for dyscalculia using the AAD model if they demonstrated a 15- point discrepancy between IQ and either the Woodcock Johnson-Revised (WJ-R) Math Calculations subtest score or the TEMA -2 Calculations subtest score. IQ scores were initially measured using the Stanford Binet - Fourth Edition (SB-IV) and later measured during year four of the study using the Wechsler Abbreviated Scale of Intelligence (WASI).

The findings of this study indicated that the prevalence of dyscalculia varied greatly depending on the identification model used. Mazzocco and Myers (2003) \[12\] stated that children identified with dyscalculia using a single assessment score (low achievement) did not necessarily demonstrate the same level of underachievement throughout four years of the study. Therefore, Mazzocco and Myers (2003) \[12\] cautioned against using scores from a single point in time to identify dyscalculia. Additionally, the authors indicated that the discrepancy model was highly inconsistent. Mazzocco and Myers (2003) \[12\] reported that some students who did not meet the discrepancy criterion had persistent, intensive mathematical difficulties, indicating a lack of sensitivity for dyscalculia identification using the AAD.
Shalev, Manor, and Gross-Tsur (2005) [14] discussed the findings of an AAD model longitudinal study. Students were identified with dyscalculia in the fifth grade, re-examined in eighth grade, and then again in eleventh grade. The researchers evaluated and monitored achievement for these students through an individually administered arithmetic battery that was standardized by "age-matched controls" (Shalev et al., 2005) [14]. Researchers identified students with dyscalculia based on an IQ of at least 80, and a score on an arithmetic battery (a neurocognitive model that evaluated number comprehension, number production, number calculation, and arithmetic error types) that was at or below the 5th percentile compared to same-grade peers. Initially, 3,029 students were assessed, and based on two follow-up assessments, 140 students were identified with dyscalculia (approximately 5 percent of the sample).

Shalev et al. (2005) [14] found that 47 percent of the students identified with dyscalculia still qualified for the disorder (based on math achievement in the lower 5th percentile of their school grade) in the eighth grade. Additionally, 95 percent of the students identified with dyscalculia were achieving in the lowest quartile of their class in the eighth grade. When the students were tested in the eleventh grade, 95 percent of the sample again scored in the lowest quartile of their class. Forty percent of the students originally identified with dyscalculia performed in the lower 5th percentile in the eleventh grade follow-up assessment.

Response to Intervention (RTI)
Fuchs et al. (2005) [3] conducted a study comparing the RTI, LA, and AAD models for identifying dyscalculia in elementary students. The authors assessed 564 first grade students using Curriculum-Based Measures (CBM) in Computation, Addition Fact Fluency, Subtraction Fact Fluency, and Concept/Applications. Low achieving students were identified based on a factor score across the CBM measures. Eleven students were additionally identified as low achieving/at-risk by participating teachers. The researchers then administered an individualized battery of tests in order to pinpoint the lowest performing students. These at-risk students were randomly assigned to a concrete-representation-abstract model tutoring group (in addition to typical classroom instruction) or a control group (typical classroom instruction).

The authors indicated that the prevalence of dyscalculia varied based on the identification model used (i.e., AAD, LA, or RTI). The results indicated that using the RTI method was a feasible approach for identifying concepts/application deficits in first grade mathematics. The authors also indicated that using achievement measures (e.g., CBMs) might result in early dyscalculia identification. Fuchs et al. (2008) concluded that RTI is an effective alternative to dyscalculia identification than the traditional AAD model.

Building on their previous 2005 study, Fuchs et al. (2007) [4] investigated the use of CBM and progress monitoring as a tool for identifying dyscalculia. The authors were interested in the predictive utility of screening measures (e.g., Number Identification/Counting, Fact Retrieval, CBM Computation, and CBM Concepts/Applications) administered in first grade (N = 170) to predict math disability at the end of second grade. A total of 27 CBM Computation and Number Identification/Counting tests were administered and recorded weekly.

Fuchs et al. (2007) [4] defined dyscalculia as performance below the 10th percentile at the end of 2nd grade on the measures of CBM Concepts/Applications and CBM Computation. Results of logistical regression analyses indicated that performance on the Concepts/Applications CBM were the best predictors of dyscalculia. Fuchs et al. (2007) [4] summarized that the Computation CBM, not the Number Identification/Counting CBM, was a valid tool for progress monitoring to verify students at risk for dyscalculia.

Low Achievement (LA) Model
Gear, Hoard, Byrd-Craven, Nugent, and Numtee (2007) [7] implemented a study which used the LA model to identify students with dyscalculia. The authors used specific cutoff scores to discriminate between students with dyscalculia, students considered to be low achieving, and typically achieving students. The participants were tested in the spring of their kindergarten year and in the fall and spring of their first grade year. The spring assessments included measures of achievement and intelligence (the authors limited this study to students with average IQ). The fall assessments included measures of processing speed and mathematical tasks.

Using cutoff scores derived from the Numerical Operations subtest of the Wechsler Individual Achievement Test – II – Abbreviated, students were diagnosed with dyscalculia if their math achievement scores were at or below the 10th percentile for at least 2 years. Students were identified as low achieving if their math achievement was between the 11th and 25th percentiles for at least 2 years. Students above the 25th percentile were identified as typically achieving. Participants in the study were administered a battery of tests measuring mathematical cognition, working memory, and processing speed as well. The children identified with dyscalculia showed deficits across all math cognition tasks, as well as working memory and processing speed. Compared with the typically achieving group, the low achieving students were less fluent in numerical processing and knew fewer addition facts. Geary et al. (2007) [7] stated that the use of a restrictive cutoff criterion may help to identify children with pervasive and severe math cognition deficits as well as underlying deficits in working memory and processing speed. The authors also reported that the use of a lenient criterion may identify children that have subtle deficits in some math domains.

Murphy, Mazzocco, Hanich, and Early (2007) [13] implemented an LA model study that examined the outcome of using different cutoff scores to identify dyscalculia in a group of kindergarteners. Specifically, the authors examined the rate of math skills growth in individuals who performed at or below the 10th percentile (n = 22), those who performed between the 11th and 25th percentiles (n = 42), and those who performed above the 25th percentile (n = 146). This study followed the participants from kindergarten through third grade, in order to include a measure of the persistence of dyscalculia, and attempted to investigate cognitive characteristics associated with math achievement.

The results of the study indicated that the participants whose performance was at or below the 10th percentile demonstrated a significantly slower rate of growth than both of the other groups. In addition, the lowest achieving (below 10th percentile) group was not only behind in initial performance measures, but also on all subsequent achievement measures into the third grade. The authors
discussed a possible plateau in achievement for the lower achieving group around 2nd grade, and that the performance of the students in the middle performing (11th – 25th percentile) group appeared to have improved over time more than the lowest group. In addition, IQ, visual–spatial ability, working memory, and rapid number naming were significant predictors of math achievement growth. The researchers stated that a possible explanation for performance differences in the two lowest performing was a deficit in a specific area(s) of math ability (e.g., number sense); however, more research is needed to verify their explanatory claim.

Watkins, Kush, and Schaefer (2002) conducted an LA model study which investigated the diagnostic utility of the Learning Disability Index (LDI), combined with the Wechsler Intelligence Scale for Children – Third Edition (WISC-III), to identify students with dyscalculia. The sample included 168 students identified with dyscalculia. The school districts which participated in the study defined dyscalculia as a discrepancy of at least 15 or more points between predicted and actual math achievement, and with no signs of a reading disability (discrepancy of less than 15 points between predicted and actual reading achievement). WISC-III assessment data (obtained within 3 years of the study) were gathered from 2,979 students enrolled in special education programs (2,274 identified as LD) within 40 Arizona school districts. Of the participants, 71.9 percent were male, ages ranged from 6 to 16 years, and grade level ranged from kindergarten through 11.

Using the provided WISC-III data, Watkins et al. (2002) calculated LDI scores using the specified methods and formula. Follow-up analyses revealed that the LDI resulted in an accurate (i.e., true positive) dyscalculia diagnosis in approximately 59 percent of cases. Based on these findings, the researchers stated that the LDI is not a reliable diagnostic indicator of dyscalculia as it can result underidentification. Watkins et al. (2002) indicated that despite these findings, and similar previous findings, many diagnosticians commonly rely on the LDI and similar diagnostic indicators for dyscalculia diagnosis.

Mazzocco, Devlin, and McKenney (2008) implemented an LA model study that compared the performance of eighth grade students with dyscalculia (n = 16), students with slightly below average math achievement (n = 19), and typically achieving students (n = 100). Students were identified as having dyscalculia if two or more of their scores on the Woodcock Johnson - Revised (WJ-R) Math Calculation subtest fell below the 10th percentile. Students were considered low achieving if two or more of their WJ-R scores fell within the 11th to 25th percentile. The authors found that performance linked to low math achievement differed from that of typically achieving student as a function of how math disabilities were defined (dyscalculia or low achieving). Additionally, the frequency and types of errors made by participants differed within the dyscalculia group, indicating heterogeneity even when a strict cutoff criterion (10th percentile) was implemented to define the group. Finally, the researchers found that the types of errors made by children with dyscalculia relied on cognitive processes other than just retrieval.

The researchers issued a cautionary statement regarding combining groups of children in order to represent the construct of dyscalculia. Similar to previously discussed studies, the heterogeneity amongst students identified as having dyscalculia was emphasized. The investigators concluded by suggesting that future studies investigate the types of errors reported in the study, which may pinpoint cognitive mechanisms underlying dyscalculia and lead to a more refined identification process.

Mazzocco (2001) conducted a study which used the LA model to identify dyscalculia in five and six year old girls with Fragile X syndrome (n = 14), Turner Syndrome (n = 9), and Neurofibromatosis Type 1 (n = 11). Participants in this study were administered a battery of tests which included: the TEMA - 2, the The KeyMath - Revised (KM - R), and the Stanford Binet – Fourth Edition (SB – IV). Students who scored at or below the 10th percentile on all measures were considered at risk for dyscalculia. The author stated that 43 percent of the girls with Turner Syndrome, 56 percent of the girls with Fragile X, and 18 percent of the girls with Neurofibromatosis were identified as at risk for dyscalculia. While the very specific population of this study is a limitation in regards to the purposes of this review, the findings do provide insight into the diagnostic utility of the LA model. For example, the author indicated the greater sensitivity of the TEMA - 2 for detecting math difficulties in kindergarten students. In sum, Mazzocco (2001) indicated the discrepancy in scores on the TEMA – 2, KM – R, and SB - IV suggests researchers need to avoid basing dyscalculia identification on single data points despite the rigorous standardization of these measures.

Stock, Desoete, and Roeyers (2010) conducted a three-year longitudinal study (n = 471) in which students were identified as having dyscalculia, persistent low achievement, or typical achievement. Initial math achievement scores were calculated using the Test for the Diagnosis of Mathematical Competencies (TEDI-MATH) at the end of participants’ kindergarten year and at the beginning of their first grade year. Participants who scored at or below the 10th percentile were diagnosed with dyscalculia, participants who scored between the 11th and 25th percentile were identified as low achieving. Follow-up math achievement scores for each student were obtained in first and second grade using the Kortrijk Arithmetic Test–Revised (KRT – R). Initial measures found that 16 of the 464 participants, (3.45 percent of the sample) scored at or below the 10th percentile for math achievement on the TEDI-MATH. Based on subsequent KRT – R achievement measures, Stock et al. (2010) reported that 87 percent of the initial 16 participants identified with dyscalculia in kindergarten were accurately identified.

The authors summarized their study by indicating the need for researchers to use restrictive cutoff scores (e.g., 10th percentile vs. 11th – 25th percentile) to distinguish between children with dyscalculia and those who are low achieving. The authors also advised researchers to refrain from making dyscalculia identification decisions until arithmetic scores from at least two consecutive years are obtained. These recommendations were made based on the results indicating the difficulty of discriminating between persistent low achievers, children with inconsistent disabilities, and/or children with dyscalculia.

Summary
This review of the literature examined the diagnostic utility of three dyscalculia identification models. Developing a reliable and valid model of identification is a crucial step in providing struggling students with the intervention they need to succeed academically, personally, and socially. Students...
with dyscalculia will face many mathematical challenges in their adult lives, so it is essential to identify those in need of the intensive, individualized instruction that special education can provide. Without a consistent model of identification, many students with dyscalculia may not be identified for the services entitled to them.

Fletcher, Fuchs, & Barnes (2007) \[5\] indicated many faults of the AAD model: a lack of external validity evidence across domains, consistent under-identification, test measurement errors that are magnified by computation of difference scores, and the similarity of individuals around the cut-points. Consistent with the findings of Fletcher et al. (2007) \[5\], the literature examined in this review indicated the low reliability of the AAD model (Fuchs et al., 2005) \[3\]. In addition, Mazzocco and Myers (2003) \[12\] reported that the AAD model leads to underidentification of students with dyscalculia. Based on these findings, and the reports of previous research, the diagnostic utility of the AAD model appears minimal.

In contrast, Fletcher et al. (2007) \[5\] described the RTI model as having strong validity (based on serial assessments and neuroimaging tests), and high reliability (multiple assessments to measure change over time). Consistent with the findings of Fletcher et al. (2007) \[5\], two of the studies reviewed above indicated that the RTI model is valid (Fuchs et al., 2005) \[3\] highly reliable (Fuchs et al., 2005; Fuchs et al., 2007) \[3, 4\], and a useful tool for progress monitoring (Fuchs et al., 2007) \[5\]. Despite the encouraging findings of the aforementioned studies, many scholars still express concerns regarding the overall diagnostic utility of the RTI model. Fletcher and Vaughn (2009) \[9\] stated that there is no consistency amongst school districts when using CBM benchmark measures to determine students’ progression through the tiers of intervention. The authors also indicated that there are no widely accepted criteria for identifying inadequate responders using the RTI model.

Fletcher et al. (2007) \[5\] described the LA model as highly valid (subgroups can be differentiated due to external variables), and of low reliability (based on decisions made from single measurements). The studies examined in this review yielded similar results. For instance, Geary et al. (2007) \[7\] indicated the LA model was a highly valid framework of dyscalculia identification. Other studies in the review found the LA model to be unreliable (Fuchs et al., 2005; Mazzocco & Myers, 2003; Watkins et al., 2002) \[5, 12, 17\]. Mazzocco et al. (2008) \[11\] indicated the LA model does not account for heterogeneity amongst students identified with dyscalculia and is consequently of limited prescriptive or diagnostic utility to special education teachers.

Based on the models examined, there seems to be apparent flaws and potential benefits of each model, suggesting a need for an alternative approach to dyslexia identification, which will be discussed in the following section.

**Future Directions**

Fletcher et al. (2007) \[5\] suggested a hybrid model of LD identification that could be useful for diagnosing students with dyscalculia. The hybrid model proposed by the authors would incorporate three primary criteria: (1) low achievement, (2) an insufficient response to scientific, research-based interventions, and (3) factors such as intellectual disability, emotional disturbance, sensory impairment, limited language proficiency, and poor instruction are ruled out. Using the hybrid model offered by Fletcher et al. (2007) \[5\], consistency for diagnosing dyscalculia would allow for greater consistency across districts and clinical settings. These scholars indicated that, through use of the hybrid model, the presence of dyscalculia could only be determined based on an achievement deficit and a documented unresponsiveness to research-based interventions. As demonstrated by the examination of studies in this review, the existing models of dyscalculia lack the validity and reliability necessary to make sound decisions regarding a student’s need for special education services.

Furthermore, future longitudinal studies should investigate the persistence of dyscalculia over extensive periods time (e.g., from kindergarten through the secondary level). Studies have indicated that characteristics of math performance are inconsistent across time and associated difficulties can manifest at different points in a person’s life (Murphy et al., 2007) \[13\]. None of the studies examined in this review spanned more than six years. Research is needed to provide a better understanding of the nature of dyscalculia over the course of multiple grade levels (> 6 years). Understanding the characteristics of dyscalculia in the context of age and developmental milestones could provide valuable information regarding identification procedures and associated interventions.

The literature also makes it apparent that further research is necessary for pinpointing the underlying cognitive mechanisms leading to dyscalculia. This obvious lack of knowledge regarding the root causes of dyscalculia seems to be a primary issue at the heart of the classification debate. How can researchers, diagnosticians, psychologists, and medical practitioners begin to identify a disability that is yet to be fully understood? It may be argued that identification can be based on the manifestation of symptoms such as poor fact retrieval, consistent calculation errors; however, studies have shown that the types of errors demonstrated by students with the most severe difficulties can overlap with the types of errors made by students whose low achievement is only marginally less than typical peers (Mazzocco et al., 2008) \[11\]. For that reason, it is apparent that researchers should incorporate the question “what is dyscalculia?” in tandem with the question “how do we diagnose dyscalculia?” in future research studies.

**References**


