Identifying Learning Disability Screens for Spanish-Speaking Adults: A Helpful Tool for Social Work and Welfare Practice

Ingrid M. Cordón
Jia Y. Chong
Kevin Grimm
Catherine Christo
Macaria Mendoza
Amanda Clinton
Gail S. Goodman

Abstract: The successful screening for possible learning disabilities (LD) is a crucial first step in the process of identifying signs of LD, gaining assistance and/or accommodations, and obtaining a more complete LD assessment. Although Latino people are the largest ethnic minority in the United States, and more specifically in California, there remains a clear need for a valid LD screening measure that is appropriate for adult Spanish speakers, particularly low-income individuals. This study evaluated the validity of three brief measures to screen for LD among low-income Spanish-speaking adults: Empire State Screen, Welfare-to-Work [WTW] 18, and MATILDA-R. The study also provides an initial estimate of LD risk in the low-income Spanish-speaking population. To estimate the predictive utility of each screening measure, 1,040 Spanish-speaking adults were administered each of the three screens and then assessed for indications of LD using multiple scoring methods (Bateria Discrepancy Diagnosis [BDD], pattern of strengths and weaknesses [PSW], and DSM-5). The translated WTW 18 Screen and the MATILDA-R appeared most promising. A culturally-sensitive, validated LD screen will help ensure that social workers and other helping professionals have access to appropriate and legally required interventions for this marginalized population.

Keywords: Learning disabilities, Spanish-speaking adults, screening, low-income, Americans with Disabilities Act

Learning disabilities (LD)--generally defined as a heterogeneous group of disorders that include substantial difficulties in such abilities as listening, speaking, reading, writing, reasoning, or mathematical abilities--are lifelong conditions that interfere with individuals' ability to understand and retain information (Swanson et al., 2013; see Individuals with Disabilities Education Act, 2004). Estimates of LD among adults are generally sparse (Gerber, 2012) and vary widely from 2% to more than 50%, depending on the segment of the population studied (e.g., adult education students, general population, prison population), the age range, and the method employed to determine the prevalence rate (Bronson et al., 2015; Corley & Taymans, 2002; Cortiella & Horowitz, 2014; Gerber, 2012).

_____

Ingrid M. Cordón, PhD, Research Scientist, Department of Psychology, University of California, Davis (UC-D), CA. Jia Y. Chong, MS, Doctoral Student, University of Illinois, Urbana-Champaign, Champaign, IL Kevin Grimm, PhD, Professor, Department of Psychology, Arizona State University, Tempe, AZ. Catherine Christo, PhD, Professor Emerita, Department of Psychology, California State University, Sacramento, CA. Macaria Mendoza, MA, Assistant Specialist, Department of Psychology, UC-D, CA. Amanda Clinton, PhD, Senior Director, Office of International Affairs, American Psychological Association, Washington, DC. Gail S. Goodman, PhD, Distinguished Professor, Department of Psychology, UC-D, CA.

Moreover, deficits in areas of functioning associated with LD (e.g., attention, working memory) persist into adulthood (Gerber, 2012; Klassen et al., 2011) and may lead individuals with LD to struggle with navigating through adult responsibilities and roles. For instance, individuals with LD have high school drop-out rates that are 2 to 3 times higher than that of their peers (Cortiella, 2013; Horowitz et al., 2017), are less likely to be enrolled in postsecondary education (though rates are increasing; Lightfoot et al., 2018; Newman et al., 2010; Wagner et al., 2005), are more likely to lag behind their peers in postsecondary completion rates (Cameto et al., 2011; Lightfoot et al., 2018), and are more likely to be unemployed than individuals in the general population (Gerber, 2012; Holliday et al., 1999; Skellern & Astbury, 2014). Due to the challenges faced by adults with LD, individuals below the poverty line were twice as likely to self-report LD than those living above poverty (Cortiella & Horowitz, 2014). Thus, providing appropriate instructional methods, programs, and accommodations may help to increase the probability of successful outcomes among LD adults.

Screening for Learning Disabilities

Successful screening for LD identifies characteristics or signs of LD through a preliminary, systematic procedure. This is a crucial first step towards obtaining a more complete assessment from a qualified professional. By undergoing screening and assessment for LD, individuals are then able to access appropriate instructional methods, programs, and accommodations that can increase successful outcomes (Gerber, 2012; Shapiro & Rich, 1999; Taymans, 2009). There is, however, a great need for the development of screening procedures for non-English speaking individuals.

Although LD occurs in diverse cultures and economic groups (Jiménez & García de la Cadena, 2007; Sideridis, 2007), and the Individuals with Disabilities Education Act (IDEA Act) gives all children regardless of culture, or economic or social disadvantage, the right to full educational opportunity, little research exists on LD among minorities and non-English speaking adults in the United States. For many young adults in the United States, identification of LD may occur during the school-age period. However, not all individuals with LD are correctly identified during childhood (Miles, 2004; Moats & Dakin, 2008. It is less likely for individuals to be identified as having LD if they were born and educated in developing countries where LD is neither fully recognized nor defined (e.g., Latin America; Agrawal et al., 2019).

Of note, individuals of Latino ethnicity represent a large part of the U.S. population (19%) and, in California, Latinos make up 39% of the state's population (U.S. Census Bureau, 2021). In 2020, Latinos comprised 36% of all Temporary Assistance for Needy Families (TANF) recipients in the U. S. and as many as 59% of all CalWORKs (California's TANF program) recipients (California Department of Social Services, 2019; U.S. DHHS, 2021). Furthermore, an estimated 8% to 18% of the adult welfare population may have learning disabilities, with some estimates as high as 40% (Dondorf-Brooks et al., 2020; Kauf, 2008; Loprest & Maag, 2001; Sweeney, 2000). Adults with LD are more likely to be sanctioned and lose their benefits than those without LD (Pavetti, 2018). Yet, a diagnosis of LD may entitle them to time-limit exceptions in TANF programs, preventing

the loss of much needed assistance (e.g., California AB-1728 CalWORKs, 2022). Given that the prevalence of LD among low-income adults is considerably higher than in the general population, many Spanish-speaking adults living in poverty in the U.S. may not have been identified as LD due to the lack of a valid LD screening measure, keeping them and their families living in poverty (Goldberg, 2002; Kusserow, 1992; Young & Browning, 2005). Thus, these individuals may not be fully benefitting from services as outlined in the Americans with Disabilities Act.

Diagnosis of Learning Disabilities

There are no federal regulations regarding the assessment measures required to identify adults with LD. Generally, assessment involves the examination of patterns of strengths and weaknesses in performance and/or achievement relative to age, as reflected across various tests, such as standardized tests of ability and achievement, and questionnaires of functioning (Gregg et al., 2006). When Public Law 94-142 was enacted in 1977 (guaranteeing free appropriate public education to children with disabilities), LD was identified by a statistically significant discrepancy between an intelligence quotient (IQ) and achievement (IQ-achievement discrepancy method; Fletcher & Miciak, 2019; Grigorenko et al., 2020). Although the Individuals with Disabilities Education Act (IDEA, 2004; Public Law 108-446) states that "a local educational agency shall not be required to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability," (Section 1414(b)(6)) historically the key concept of a discrepancy between intellectual potential and academic performance or achievement continued to be used in the assessment process (e.g., Fletcher & Miciak, 2017; Gregg et al., 1999; Seo et al., 2008). A non-severe IQ/achievement discrepancy, for instance, can be part of the LD identification process for vocational rehabilitation services (U.S. Department of Education, 2022). Research on individuals with LD often includes standardized measures of intelligence and achievement (e.g., Wechsler Intelligence Scale for Children, Stanford-Binet, Woodcock-Johnson Tests of Cognitive Abilities) as part of the assessment procedure (e.g., Johnson et al., 2010; Williams et al., 2016).

In addition to the pattern of profiles on standardized tests, a key factor in the accurate assessment of LD involves the judgment of a trained clinician (Fiorello et al., 2014; Fletcher et al., 2018; Franz et al., 2017; Maki et al., 2015; Miciak & Fletcher, 2019), who considers the environmental, biological, cognitive, language, and behavioral factors influencing an individual's ability to learn tasks in a specific context (Fiorello et al., 2014). Overall, when determining the presence of a learning disability, clinicians must consider the interaction of multiple factors that impede learning, foster learning, or mitigate the influence of those factors that impede learning. Clinicians would consider, for example, that individuals with similar cognitive profiles might have different levels of academic success due to factors such as supportive environments, behavioral characteristics, and personality traits. Gregg and colleagues (2006) argue that a balance between statistical data (as provided by standardized measures) and clinical judgment is needed to assess LD (but see Fletcher & Miciak, 2019; Miciak & Fletcher, 2020).

It is important to note that, during the course of this study, the American Psychiatric Association (APA, 2013) revised the criteria for determining LD, as delineated in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5). The new criteria place less emphasis on discrepancies in IQ/Achievement than did the previous version (i.e., DSM-4) and require evidence of a childhood history of LD. We anticipated that this change in the DSM criteria would reduce the number of adults, specifically Spanish-speaking adults, who might have been diagnosed with LD, given that most of our participants were from Latin America (primarily Mexico) where LD is not fully recognized and/or screened for in childhood (Agrawal et al., 2019; Stough & Aguirre-Roy, 1997). To address the change in criteria established by the 5th edition of the DSM, both the DSM-4 and DSM-5 criteria were used in the present study to determine the clinical LD status of participants and to evaluate the pilot LD screens.

Overview of the Present Study

We were asked by the California Department of Social Services (CDSS) to test various LD screening measures that could be used with Spanish-speaking CalWORKs (California TANF) clients. A Spanish-language screening measure would allow CDSS to identify clients needing further LD assessment and thereby identify clients who could benefit from additional CDSS services (see also Holcomb & Thompson, 2000). The current study, therefore, evaluated the predictive utility and validity of three pilot screens of LD among low-income adult Spanish-speakers, specifically those with CalWORKs eligibility. Moreover, the study provides an initial estimate of the prevalence of LD risk in this population. To increase the likelihood of finding a sufficiently large subsample of Spanish-speaking LD adults, recruitment focused on individuals who were likely to meet California's welfare (i.e., CalWORKs) eligibility requirements (i.e., 18 years of age or older, low family monthly income, being pregnant with or the caretaker of one or more children).

Three screens were selected from the existing literature: the Empire State Screen, the Welfare-to-Work (WTW) 18 CalWORKs LD Screen, and the Mississippi Assessment Technique for Identifying Learning Disabilities in Adults (MATILDA-R). These screens were chosen based on their statistical properties, brevity, and ease of administration, scoring, and interpretation.

Tests of intellectual ability and achievement are often still core components of assessing LD. Thus, eligible participants for our study were administered two standardized assessments measuring general intellectual ability (Bateria III and Test of Nonverbal Intelligence, 4th edition [TONI-4]) to determine each participant's LD status and the accuracy and predictive utility of the screens within this sample. In addition, a clinical LD specialist diagnosed each participant using two different methods (the patterns of strengths and weaknesses [PSW] and DSM-5 approaches). The overall aim of the study was to identify a culturally-sensitive, validated LD screen that would help social workers and others ensure legally required interventions. In the following sections, the term *significance* refers to statistical significance.

Methods

Participants

A total of 1,107 Spanish-speaking, low-income adults participated in the current study (884 females; Mage = 36 yrs, SD = 8.34). Data collection occurred between January, 2012 and January, 2014. Data for 67 participants were not included for the following reasons: substantial disruptions during testing (e.g., distracting environmental noise), participant illness during testing, incomplete tests, errors in testing administration, and participants who were found ineligible after the testing session. Participation criteria were as follows: 1) age 18 years or older; 2) parent of at least one child or currently pregnant; 3) if married/cohabitating, partner was unemployed or worked less than 100 hours per month; 4) renter (not homeowner); 5) Spanish as their native/dominant language; and 6) low-income status. To ensure the participants were of sufficiently low income, these criteria, as well as the testing of a predominantly female sample, were used due to being consistent with CalWORKs eligibility determination at the time.

The majority of the participants were born outside of the United States (97%), primarily in Mexico (89%). On average, participants were educated up to an 8th grade level in a non-U.S. country. Most of the participants also indicated that they did not speak or write English well (81%). About half of the sample was married (54%), with a mean family income of $851 per month and an average of 2.83 children per household. Seventy-two percent of the participants were unemployed, and more than half of the participants reported receiving some form of public assistance. All procedures were approved by the University of California, Davis' Institutional Review Board. Written informed consent in Spanish was obtained from all participants, and they were compensated for their time.

Materials

All study materials were either originally developed in or translated into Spanish for this study. Materials included a demographic questionnaire to determine eligibility for enrollment and to obtain background information, three pilot screening measures (Empire State Screen, WTW 18, and MATILDA-R), and two standardized measures of general intellectual ability (Bateria III and TONI-4). To ensure the accuracy and appropriateness of the translated screens, a focus group of fluent Spanish speakers was first conducted. Changes to the language of the screen were made based on focus group feedback and finalized by native Spanish-speaking postgraduates. In addition, each pilot screen was back-translated to verify the reliability of the translation.

Demographic (eligibility) questionnaire. A demographic questionnaire was developed to obtain background information and determine participant eligibility. Background questions covered, for example, age, birthplace, marital status, highest education, and number of children.

Empire State Screen. The Empire State Screen is a composite of 11 items selected for their statistical diagnostic utility from four candidate screening measures (Abwender, 2005). The screen was written in Spanish and tested on Spanish-speaking, low-income adults. Because the Empire State Screen measures both learning disability and marginal intellectual functioning, it was not designed to differentially diagnose LD. Therefore, the final screen is more accurately described as identifying "learning needs" as opposed to specifically being a screen for LD.

The Empire State Screen is comprised of "yes/no" statements concerning learning-related problems. For scoring, the weighted point values associated with each participant's responses are summed and a constant value (i.e., 614) is subtracted from the sum. Scores above 50 after subtracting the constant value are considered to reflect the presence of learning needs. According to Abwender (2005), the Empire State Screen is 83% accurate (83% sensitivity, 84% specificity) in the identification of learning needs in adults. Reliability for our sample was $\alpha = .72$.

Welfare-to-Work (WTW) 18. This tool is based on the Washington State Screen (WSS) developed by Payne & Associates in collaboration with the Washington State Department of Social and Health Services, to identify learning needs in low-income English speakers (DSHS, 1998). The WSS was translated into Spanish (WTW 18) for the present study. The original screen consists of 13 "yes/no" questions regarding learning-related problems that were divided into four differentially weighted sections. Scores at or above 12 are considered to reflect a high risk of LD. The WSS has an overall diagnostic accuracy of 74% (70% sensitivity, 79% specificity) for learning needs within an English-speaking population (DSHS, 1998). Although the reliability for the WTW 18 was only .42 overall, fortunately, reliability for a key WTW 18 index (13 items from the WTW, unweighted), called the Washington Unweighted Score (WUS), was .76 for our sample, and thus an acceptable score. The WTW was therefore included in the analyses, but with a focus on the WUS index.

MATILDA-R. The Mississippi Assessment Technique for Identifying Learning Disabilities in Adults (MATILDA; Dorsett, 2000; Grubb et al., 1997; Lancaster, 2002) consists of both "yes/no" questions and several tasks (e.g., writing the numbers 1-20). Scores at or above 13 suggest a risk of LD (see MATILDA scoring guidelines; Grubb et al., 2001).

The MATILDA, developed for use with college-level, English-speaking students, was revised for our study (the MATILDA-R) to better reflect the educational level, language, and culture of the target population. Modifications to the MATILDA included changes to the instructions (e.g., "write the Spanish alphabet"), an additional example in the organizational skills section, shortening the paragraph participants were asked to reproduce, more culturally representative protagonist names, and simpler math problems. The screen was then translated into Spanish.

To ensure the accuracy and appropriateness of the translated screen, it was presented to a focus group of Spanish-speaking participants for review and feedback. Following the translation, the revised screen was tested on a small group ($n = 10$) of bilingual (English and Spanish-speaking) participants to test the face validity of the revised screen against the

original screen. Reliability for the MATILDA-R ranged from .20 to .73 across sections. Permission to revise, translate, and use the modified screen was obtained from the lead researcher, Dr. Grubb.

Woodcock-Johnson Bateria III. The Bateria III, developed for Spanish-speakers who are 2 to 90+ years of age, is a comprehensive set of tests that measures both cognitive abilities and achievement levels. The Bateria III interpretive plan is based on cluster (grouped items) interpretation. Most of the Bateria III subtests and clusters show reliabilities at .73 or higher (Muñoz-Sandoval et al., 2005).

The Bateria III also provides two major types of discrepancy scores: ability/achievement discrepancies and intra-achievement discrepancies. The ability/achievement discrepancy is the most commonly used method of diagnosing LD. Generally, LD is indicated if there is a difference of at least 15 points (≥ 1.5 SD) between the ability and achievement subscales, which are based on comparisons to age-normed groups upon which the Bateria III was developed. Information gathered from intra-ability discrepancies help professionals determine an individual's strengths and weaknesses and diagnose and document language and learning disabilities.

Test of Nonverbal Intelligence, 4th edition (TONI-4). The TONI-4 assesses general intellectual ability (abstract reasoning, figural problem-solving) without allowing factors such as poor language or lack of cultural knowledge to conceal an individual's intelligence. The TONI-4 showed a reliability of $\alpha = .92$ in our sample. Convergence in the estimates of intellectual ability of the Bateria III and the TONI-4 suggest that the Bateria III, despite its greater reliance on language, provides a relatively accurate representation of participants' intellectual ability.

Procedure

Recruitment. Participants were recruited from multiple agencies in California, across 13 west coast counties, that provide services to Spanish-speaking, low-income adults, including community health clinics, Head Start centers, adult education schools, family resource centers, and county social services offices. Potential participants were approached by trained Spanish-speaking research assistants (RAs) at the various recruitment locations. The RAs briefly explained the study to potential participants as they waited for or left appointments. Interested persons provided a number for call-back and/or were given a flyer with our contact information. Although the vast majority of participants (approximately 90%) were actively recruited into the study, additional participants were added using a snowball sampling method (a non-probability method used to locate hard-to-reach populations; Johnson, 2014) as a means of recruiting this hard-to-reach population. All potential participants were contacted by phone and administered a demographic questionnaire. If interested and eligible, an appointment for testing was scheduled. At the end of the session, participants received a $150 gift card.

Screen and assessment administration. The RAs administering the screens and assessments received an intensive, week-long training with a certified Bateria III trainer who provided training on the administration, scoring, and interpretation of the Bateria III

and modeled all testing procedures. The RAs practiced with the trainer and with each other. Finally, the RAs were given additional practice and training by the Co-Principal Investigator, a highly experienced psychologist in testing and assessment (interrater reliability $\kappa = 0.76$, $p < 0.001$). Fidelity of all testing measures and procedures, in terms of adherence, exposure/duration, and quality of delivery, was established for each RA prior to permitting the RA to collect data. The fidelity of the screens and assessments were also monitored throughout the project.

Participants were tested individually in Spanish in a quiet room by two graduate-level, Spanish-speaking RAs. Because the full administration took about 4 hours, participants were given breaks to minimize fatigue. The three screens were administered first by one of the RAs, with screen order counter-balanced across participants. Following the screen administration, a different RA who was unaware of the screen results administered the two counterbalanced full-battery intelligence assessments. Materials were read to participants in Spanish.

Screen scoring procedures. The screens were scored in accordance with guidelines provided by the test developers. The screens were selected to be appropriate for low-income Spanish speakers and to be easily scored, requiring only 0 (incorrect) or 1 (correct) scoring per participant answer. To increase flexibility in examining various scoring methods that could potentially maximize the screen's possible predictive function, two or more scores were derived for each screen. These scores included the recommended clinical guidelines provided with each screen and one or more total raw scores.

Empire State Screen scoring. Two scores were derived from the Empire State Screen data, the Empire State Diagnosis (ESD) and Empire Total Score (ETS). The ESD was derived from the total number of weighted "yes" responses. A recommended clinical cut-off score of 51+ points (Abwender, 2005) was employed to determine the predicted LD status. The ETS is the sum of weighted "yes" responses plus the sum of weighted "no" responses, minus 614.

WTW 18 Screen scoring. Three scores were derived from the WTW 18 Screen data: the Washington State Diagnosis (WSD), the Washington Weighted Score (WWS), and the Washington Unweighted Score (WUS). The WSD is derived from the weighted total number of "yes" responses (DSHS, 1998). Accordingly, a weighted score of 12 or more is indicative of LD. The WWS is the total number of "yes" responses to the four weighted scoring sections. The WUS is the total number of "yes" responses to the 13 questions (unweighted).

MATILDA-R scoring. Four scores were derived from the MATILDA-R data. The MATILDA-R Diagnosis (MRD) concurs with the English MATILDA screen guidelines (Grubb et al., 2001) and is derived from the total number of "yes" responses. An MRD score of 13 or more is indicative of LD risk. MATILDA-R Total Yes Responses (MTYR) is the total number of "yes" responses to the scoring form of this study. MATILDA-R Total Errors (MTE) is the total sum of participants' errors on the MATILDA-R. MATILDA-R Error Diagnosis (MED) was derived from the MTE scores; a clinical cut-off of 13 or more errors was used to derive the MED. The 13+ error cut-off was based on preliminary analyses of the MATILDA-R, which suggested a significant association between the MED

(on the one hand) and the BDD and PSW diagnoses (on the other hand) when using the 13+ cut-off score.

Scoring the Bateria-III and TONI-4. For the Bateria-III and TONI-4, each test question is scored as 0 = incorrect and 1 = correct, and the scoring was conducted during testing. The correct answers are in the scoring booklets and only require that the tester enter 0s and 1s. At the end of each testing section, the 0s and 1s were summed.

Diagnostic methods. Historically, a key component of making a diagnosis of LD is the discrepancy between intellectual potential (ability) and academic performance (achievement; e.g., Seo et al., 2008). Although no federal regulations exist for mandatory assessment measures for identifying adults with LD, assessment of adults in clinics has generally followed the DSM-4-TR (American Psychiatric Association, 2000) recommendation of examining patterns of strengths and weaknesses in performance and/or achievement in relation to age and educational level as reflected across various tests (Taymans, 2012). Standardized measures of intelligence and achievement are often included in assessment procedures in LD research across all age groups (Fletcher & Miciak, 2019; Holliday et al., 1999; Seo et al., 2008).

Change in diagnostic criteria. During the latter half of the current study, the American Psychiatric Association (APA, 2013) released the 5th edition of the DSM, which changed the classification criteria for determining LD. Included in the DSM-5 is a change in category from "Learning Disability" to "Specific Learning Disorder," or SLD. The new criteria place less emphasis on discrepancies in IQ/Achievement than the DSM-4 version and require evidence of a childhood history of LD. Because the choice regarding the most appropriate screen might depend on the diagnostic procedures used in the field, we chose to use three scoring procedures (Bateria Discrepancy Diagnosis [BDD], clinical diagnosis, and patterns of strengths and weaknesses [PSW] approach) to assess the diagnostic utility of each screen. These scoring procedures reflected both the DSM-4-TR and DSM-5 criteria, as well as a standard clinical approach.

Bateria Discrepancy Diagnosis (BDD). Raw scores from each of the Bateria III subscales were entered into a standard Bateria III software program. The software produces a summary report of standardized scores, discrepancies, and significance probabilities, including: 1) General Intellectual Ability (GIA), which represents a measure of IQ; 2) Achievement, which represents level of academic achievement; 3) standardized scores on the cognitive subtests and cluster that combine to form the GIA score; 4) standardized scores on the achievement subtests and cluster that combine to form the Achievement score; 5) intra-cognitive and intra-achievement discrepancy scores; and 6) ability/achievement discrepancies, which provide information regarding discrepancies between the various cognitive and achievement subscales and, if sufficiently large, are indicative of the specific type of LD the participant may have.

Based on the information computed by the Bateria III software, the BDD was created by examining whether any significant ($\geq$ 1.5 *SD*) discrepancy among the ability/achievement subscales was observed. The Bateria III software generates eight ability/achievement subscale comparisons (e.g., ability vs. math achievement). In the present study, a significant discrepancy among any of the eight comparisons was

considered indicative of LD.

Clinical diagnosis. A clinical specialist received the following de-identified information: 1) summary reports computed by the Bateria III software, 2) TONI-4 results, 3) scanned copies of the Bateria III and TONI-4 response form, and 4) demographic information. Using these materials, the LD specialist provided a clinical diagnosis and a brief clinical report for each participant. Due to change in DSM criteria, the LD specialist employed two methods: the PSW and DSM-5 approaches (described below). Two dichotomous (0 = not LD, 1 = LD) decision scales were developed for the PSW and DSM-5 diagnoses for each participant's LD status.

Patterns of Strengths and Weaknesses (PSW) approach. The PSW approach is based on research recommendations, guidelines for the Bateria III, and consistency with DSM-4-TR criteria. First, it is determined if the participant is performing significantly (i.e., 1 *SD*) below peers in an academic area. This is accomplished by examining the broad academic scores and selected academic clusters to see if any academic area is below the 15th percentile or a standardized score of 85. Academic sub-skills are also reviewed. If a participant is performing below the average range in an academic area, the protocol is reviewed to determine areas of weakness. In addition, the level of academic performance is compared to the educational level to see if it is substantially lower than expected given the education level attained.

After reviewing academic performance and contextual factors, the cognitive scores are examined to determine if any cognitive processes are significantly below peers in expectations for age and education level. If so, the intra-individual pattern of scores is analyzed to determine if a substantial difference exists between an area of weakness and other cognitive processes. Response patterns are also examined. The pattern of cognitive processes is reviewed for evidence of a cognitive weakness that could contribute to LD. If the participant shows an area of significant cognitive weakness, that cognitive score is compared to the area of academic weakness to determine if the two areas of weakness are known to be related. If so, then an overall analysis of the academic and cognitive scores in the context of educational, experiential, and employment factors is undertaken to determine if the profile fits a pattern of strengths and weaknesses indicative of a specific LD.

The PSW method considers all assessment data available. It relies on clinical appraisal and reflects the current method of considering the entire pattern of performance to determine LD. To help ensure diagnostic accuracy, 10% of the sample was randomly selected for a second clinical LD specialist to independently diagnose using the PSW approach. The diagnoses of the two clinicians were significantly correlated ($r = .75$, $p < .001$). Moreover, inter-rater reliability statistics indicated that the two clinicians were reliable in their coding (ICC = .86, $p < .001$). Disagreements were discussed and resolved. One of the clinicians then coded the rest of the data.

DSM-5 approach. The main clinical specialist also provided a diagnosis based on the DSM-5 criteria. The DSM-5 method used information obtained from the standardized tests, demographic information, and participants' responses to two questions on the WTW 18 Screen regarding learning problems in primary and secondary school.

To help ensure diagnostic accuracy, 10% of the sample was randomly selected for the second clinical LD specialist to independently diagnose using the DSM-5 method. The diagnoses of the two clinicians were significantly correlated ($r = .66$, $p < .001$). Inter-rater reliability statistics indicated that the two clinicians' codings were reliable (ICC = .79, $p <$ 001).

Results

The Bateria III and TONI-4 use a standardized score of 100 (+/- 15 points) as representing an average IQ. Based on the Bateria III GIA score, the mean IQ for participants was below average (78.48, $SD = 10.55$). Scores on the achievement portion of the Bateria III suggested that the mean level of achievement was within the average range of intellectual achievement (88.98, $SD = 8.68$). The mean IQ score derived from the TONI-4 was 8 points higher (85.74, $SD = 6.75$) than the Bateria GIA score, indicating an IQ within the low average range, closer to the participants' achievement scores on the Bateria III. A paired t-test indicated that the difference in IQ scores was significant, such that the Bateria III IQ score was significantly lower than the TONI-4 IQ score, $t(1039) =$ 26.91, $p < .001$.

To develop the predictive models of LD for each screening measure within this sample, univariate logistic regression was used to predict each case of LD from the scores and diagnoses available from each screen test. For each logistic regression, one of the scoring variables (e.g., ESD, ETS, WUS) was entered as a predictor of each of the LD diagnosis methods (BDD, PSW, DSM-5). Tables 1 through 5 provide descriptive statistics, correlations, and a summary of each screen's overall accuracy, sensitivity, and specificity rates across the three methods of determining LD. Estimates of LD risk is provided per screen and per assessment (Table 2).

Table 1. *Correlations Among LD Status Criteria and Pilot Screen Scoring Methods (n = 1,040)*

	BDD	PSW	DSM-5	ESD	ETS	WSD	WWS	WUS	MRD	MTYR	MED	MTE
BDD	1											
PSW	.343**	1										
DSM-5	.181**	.382**	1									
ESD	.248**	.150**	.169**	1								
ETS	.232**	.153**	.140**	.827**	1							
WSD	.188**	.050	.194**	.354**	.312**	1						
WWS	.244**	.066*	.232**	.446**	.414**	.832**	1					
WUS	.247**	.071*	.271**	.463**	.428**	.800**	.969**	1				
MRD	.329**	.265**	.199**	.353**	.327**	.314**	.379**	.389**	1			
MTYR	.367**	.257**	.174**	.420**	.388**	.385**	.479**	.486**	.820**	1		
MED	.285**	.270**	.092**	.302**	.282**	.258**	.313**	.316**	.660**	.683**	1	
MTE	.284**	.187**	.047	.286**	.275**	.267**	.323**	.316**	.596**	.739**	.735**	1

Note. * $p < 0.05$ ** $p < 0.001$. BDD = Bateria Discrepancy Diagnosis. PSW = Pattern of Strengths & Weaknesses DSM = Diagnostic & Statistical Manual. ESD = Empire State Diagnosis. ETS = Empire Total Score. WSD = Washington State Diagnosis. WWS = Washington Weighted Score. WUS = Washington Unweighted Score. MRD = MATILDA-R Diagnosis. MTYR = MATILDA-R Total Yes Responses. MED = MATILDA-R Error Diagnosis. MTE = MATILDA-R Total Errors.

Table 2. *Number of Participants Identified as Potentially Learning Disabled (n = 1,040)*

Diagnosis	Criteria based on	Screened for LD [n (%)]	
		Negative	Positive
Empire State	Scoring guidelines provided by Empire State Screen (Abwender, 2005).	734 (70.6%)	306 (29.4%)
Washington State	Scoring guidelines provided by Washington State Screen (DSHS, 1998).	888 (85.4%)	152 (14.6%)
MATILDA-R	Scoring guidelines provided by MATILDA (Grubb et al., 2001).	661 (63.6%)	379 (36.4%)
		Identified [n (%)]	
		Not LD	LD
Bateria Discrepancy	Bateria III ability/achievement discrepancy scores computed by Bateria III software.	806 (77.5%)	234 (22.5%)
PSW	Clinical diagnosis using PSW method.	868 (83.5%)	172 (16.5%)
DSM-5	Clinical diagnosis using DSM-5 criteria.	917 (88.2%)	123 (11.8%)

Empire State Screen Results

The ESD resulted in 306 of the 1,040 participants being recommended for further testing for LD. This identification of LD was significantly associated with the BDD, the PSW diagnosis, and the DSM-5 diagnosis. The ETS was examined to determine if there was a more optimal cut-off score in our population. This scoring variable (M = 57.32, SD = 101.55) explained 7% of the variability in the BDD and had a peak accuracy of 77.7%; however, this cut-off score had poor sensitivity (2.6%) as most of the accuracy was based on its specificity (99.5%). If we look to obtain a higher level of sensitivity, there is a cut-off (i.e., 10) that provides an overall accuracy rate of 67.0%, with 58.5% sensitivity and 69.5% specificity. The ETS was also significantly associated with the PSW diagnosis and the DSM-5 diagnosis. The ETS explained 3% of the variability in the PSW diagnosis and 3% of the variation in the DSM-5 diagnosis. The accuracy of the ETS is maximized when no participants are recommended for further testing; however, such a decision has 0% sensitivity. There was no "good" cut-off based on the PSW or DSM-5 diagnosis because the ETS explained little variance in these diagnoses.

Overall, the ETS and ESD were significantly associated with the BDD, PSW diagnosis, and DSM-5 diagnosis. However, the Empire State scores had poor sensitivity ratings and explained less variability in these diagnoses compared to the other two screens.

WTW 18 State Screen Results

The WSD resulted in 152 adults recommended for further testing for possible LD. This diagnosis was significantly associated with the BDD and the DSM-5 diagnosis but not with the PSW diagnosis. The WWS (M = 4.89, SD = 6.07) explained 8% of the variability in the BDD and had a peak accuracy of 78.3%; however, this cut-off score had poor sensitivity (11.5%) and its accuracy was mostly based on its specificity (97.6%). If we look to obtain a reasonable level of sensitivity, a cut-off score of 3 provides an overall accuracy rate of 61.7%, with 63.2% sensitivity and 61.3% specificity. The WWS was also significantly associated with the PSW diagnosis and the DSM-5 diagnosis. The WWS explained 1% of

the variability in the PSW diagnosis and 9% of the variation in the DSM-5 diagnosis. The accuracy of the WWS is maximized when no participants are recommended for further testing; however, such a decision has 0% sensitivity. There was no "good" cut-off based on the PSW diagnosis because the WWS explained little variance in this diagnosis. Cut-off scores either had high accuracy (i.e., > 70%) and little sensitivity (i.e., < 30%) or low accuracy and high sensitivity. For the DSM-5 diagnosis, there were cut-off scores that were fairly accurate with adequate sensitivity and specificity. A cut-off score of 6 yields an accuracy of 70.3%, a sensitivity of 63.4%, and a specificity of 71.2%.

The WUS ($M = 2.40$, $SD = 2.67$) explained 8% of the variability in the BDD and had a peak accuracy of 78.0%. However, this cut-off score had poor sensitivity (5.6%) as most of the accuracy was based on its specificity (99.0%). If we look to obtain a reasonable level of sensitivity, there is a cut-off score (i.e., 2) that is 59.3% accurate, with 66.7% sensitivity and 57.2% specificity. The WUS was also significantly associated with the PSW diagnosis and the DSM-5 diagnosis. The WUS explained 1% of the variability in the PSW diagnosis and 12% of the variation in the DSM-5 diagnosis. The accuracy of the WUS is maximized when no participants are recommended for further testing; however, such a decision has 0% sensitivity. There was no "good" cut-off based on the PSW diagnosis, because the WUS explained little variance in this diagnosis. For the DSM-5 diagnosis, however, there were cut-off scores that were fairly accurate with adequate sensitivity and specificity. An example cut-off score of 3 had an accuracy of 68.8%, a sensitivity of 73.2%, and a specificity of 68.3%.

Overall, the results from the WTW 18 Screen showed that the scores it produced were statistically significant predictors of having LD based on the BDD, the DSM-5 diagnosis, and to a lesser extent the PSW diagnosis. The WUS had the most predictive power of the WTW 18 Screen scores. This score had lower predictive power than the MATILDA-R diagnosis when the BDD and the PSW diagnosis were the outcomes of interest but had greater predictive accuracy for the DSM-5 diagnosis. Our recommended cut-off on the WUS is a score of 3 or more. This cut-off score explained 14% of the variation in DSM-5 diagnosis and had an accuracy of 68.8%, a sensitivity of 73.2%, and a specificity of 68.3%.

MATILDA-R Results

The MRD is based on the MTYR, with a cut-off score of 13 or more as indicative of a risk for LD. Based on the MRD, 379 participants would be recommended for further LD testing. This diagnosis was significantly associated with the BDD, the PSW diagnosis, and the DSM-5 diagnosis (Tables 3 to 5). Because the recommended cut-off score of 13 may not be optimal for our population, the MTYR was also examined. The MTYR ($M = 11.71$, $SD = 4.32$) explained 18% of the variability in the BDD and had a peak accuracy of 79.6%; however, this cut-off score, which optimizes accuracy, had poor sensitivity (18.8%) as most of the accuracy was based on its specificity (97.3%). For a more reasonable level of sensitivity, there is a cut-off (i.e., 12) that is 65.6% accurate, with 75.2% sensitivity and 62.8% specificity (in addition to the cut-off score that was used for the MRD). The MTYR was also significantly associated with the PSW and DSM-5 diagnoses (Tables 4 to 5). The MTYR's accuracy is maximized when no participants are recommended for further testing;

however, such a decision has 0% sensitivity. When an acceptable level of sensitivity is obtained (e.g., 60%), accuracy is 68.6% for the PSW diagnosis and 66.5% for the DSM-5 (using a cut-off score of 13).

The MED score indicated that 334 participants should be recommended for further LD testing. This diagnosis was significantly associated with the BDD, the PSW diagnosis, and the DSM-5 diagnosis (Tables 3 to 5). The MED score has a cut-off of 13 and is derived from the MTE. Because this cut-off score may not be optimal, we examined the MTE as a scoring option. The MTE ($M = 12.66$, $SD = 12.57$) explained 10% of the variability in the BDD and had a peak accuracy of 78.8%; however, this cut-off score, which optimizes accuracy, had poor sensitivity (9.8%) as most of the accuracy was based on its specificity (98.8%). For a reasonable level of sensitivity, there is a cut-off (i.e., 9) that is 62.2% accurate, with 74.8% sensitivity and 58.6% specificity (in addition to the cut-off that was used for the MED). The MTE was significantly associated with the PSW diagnosis, but not with the DSM-5 diagnosis. The MTE explained 5% of the variability in the PSW diagnosis and 0% of the variability in the DSM-5 diagnosis. The accuracy of the MTE is maximized when no participants are recommended for further testing; however, this decision would result in 0% sensitivity. When an adequate level of sensitivity is obtained (e.g., 60%), accuracy is 66.3% for the PSW diagnosis. We do not report accuracy for the DSM-5 diagnosis because the Matilda Error Score was not significantly associated with it.

Overall, the results from the MATILDA-R showed that the MRD was a significant predictor of having LD based on the BDD, the PSW diagnosis, and DSM-5 diagnosis, and that this measure (along with the MTE) was more predictive than the MED or the MTYR. When attempting to optimize the cut-off score on the MTYR, we obtained similar results to the MED. The MRD (rather than the other three MATILDA-R indices) would be the best MATILDA index for making recommendations for further LD testing, at least within this sample.

Table 3. *Bateria Discrepancy Diagnosis (BDD): Screen Accuracy, Sensitivity, Specificity, and Variance Explained*

	Accuracy	Sensitivity	Specificity	Variance Explained
Empire State Screen				
Empire State Diagnosis (ESD)	70.8%	50.4%	76.7%	9.0%
Empire Total Score (ETS)	77.7%	2.6%	99.5%	7.0%
WTW 18 Screen				
Washington State Diagnosis (WSD)	75.0%	26.9%	89.0%	5.0%
Washington Weighted Score (WWS)	78.3%	11.5%	97.6%	8.0%
Washington Unweighted Score (WUS)	78.0%	5.6%	99.0%	8.0%
MATILDA-R				
MATILDA-R Diagnosis (MRD)	70.7%	65.8%	72.1%	15.0%
MATILDA-R Total Yes Responses (MTYR)	79.6%	18.8%	97.3%	18.0%
MATILDA-R Error Diagnosis (MED)	71.0%	56.8%	75.1%	11.0%
MATILDA-R Total Errors (MTE)	78.8%	9.8%	98.8%	10.0%

Table 4. *Pattern of Weaknesses and Strengths (PSW) Clinical Diagnosis: Screen Accuracy, Sensitivity, Specificity, and Variance Explained*

	Accuracy	Sensitivity	Specificity	Variance Explained
Empire State Screen				
Empire State Diagnosis (ESD)	68.8%	44.8%	73.6%	4.0%
Empire Total Score (ETS) [1]	--	--	--	3.0%
WTW 18 Screen				
Washington State Diagnosis (WSD) [2]	--	--	--	0.0%
Washington Weighted Score (WWS) [1]	--	--	--	1.0%
Washington Unweighted Score (WUS) [1]	--	--	--	1.0%
MATILDA-R				
MATILDA-R Diagnosis (MRD)	68.6%	65.1%	69.2%	11.0%
MATILDA-R Total Yes Responses (MTYR)	60.0%	68.6%	51.4%	10.0%
MATILDA-R Error Diagnosis (MED)	71.3%	60.5%	73.5%	11.0%
MATILDA-R Total Errors (MTE)	60.0%	66.3%	53.7%	5.0%

Notes. [1] Variance explained was too low to permit the selection of a cut-off score that would allow reasonable calculation of accuracy rates. [2] Association between the WSD and PWS was not significant

Table 5. *DSM-5 Clinical Diagnosis: Screen Accuracy, Sensitivity, Specificity, and Variance Explained*

	Accuracy	Sensitivity	Specificity	Variance Explained
Empire State Screen				
Empire State Diagnosis (ESD)	70.7%	50.4%	73.4%	5.0%
Empire Total Score (ETS) [1]	--	--	--	3.0%
WTW 18 Screen				
Washington State Diagnosis (WSD)	81.4%	33.3%	87.9%	6.0%
Washington Weighted Score (WWS)	70.3%	63.4%	71.2%	9.0%
Washington Unweighted Score (WUS)	68.8%	73.2%	68.3%	14.0%
MATILDA-R				
MATILDA-R Diagnosis (MRD)	66.5%	62.6%	67.1%	7.0%
MATILDA-R Total Yes Responses (MTYR)	60.0%	66.5%	53.5%	5.0%
MATILDA-R Error Diagnosis (MED)	66.4%	43.9%	69.5%	2.0%
MATILDA-R Total Errors (MTE) [2]	--	--	--	0.0%

Notes. [1] Variance explained was too low to permit the selection of a cut-off score that would allow reasonable calculation of accuracy rates. [2] Association between the MTE and DSM-5 diagnosis was not significant.

Important Considerations in Selecting a Screening Measure

Overall, the three LD screening measures provided the highest percentage of explained variance in the BDD relative to the PSW and DSM-5 diagnoses. This is partly because the BDD had a higher base rate for LD. Thus, the greatest accuracy was achieved when using

this outcome measure. The MATILDA-R was associated with all three LD diagnosis methods: It was the best predictor of the BDD and the PSW diagnosis. The previously recommended cut-off score on the MTYR was useful for our population. The WTW 18 Screen was not as useful as the MATILDA-R when the outcome of interest was the BDD or the PSW diagnosis; the WUS from the WTW 18 Screen was most predictive of the DSM-5 diagnosis. Therefore, the screen selected depends on which LD diagnosis method is most appropriate. If the BDD or the PSW diagnosis is used, the MATILDA-R (using MRD scoring) is the best option. If the DSM-5 diagnosis is selected, the WTW 18 Screen (using the WUS with cut-off of 3 or more for scoring) may be more useful.

Of note, determining the best cut-off scores can be chosen based on objective (e.g., mathematical approaches; Dunstan & Scott, 2019) and/or subjective factors (Dunstan & Scott, 2020). For instance, recommended cut-off scores can be based objectively on the highest accuracy without consideration for sensitivity and specificity, or subjectively giving more prominence to sensitivity or specificity to meet an overarching goal (e.g., keeping false positives low; Bujang & Adnan, 2016; Dziak et al., 2020). The cut-off scores in the present study were data-driven consistent with quantitative approaches in screen and diagnostic development (Bujang & Adnan, 2016). We place a good bit of importance on correctly classifying participants who are LD, while acknowledging practical, clinical, and cultural considerations in determining cut-off scores (e.g., costs of further assessment, overdiagnosis, disproportionality of diagnosis; Shifrer et al., 2011; Zimmerman & Galione, 2011). We thus suggest trying to balance accuracy with sensitivity and specificity while obtaining an adequate level of sensitivity where possible.

Discussion

We tested three screens for LD in Spanish-speaking, low-income adults. Our findings suggest that the WTW 18 Screen and MATILDA-R are viable options for identifying low-income Spanish speakers who may benefit from a clinical LD assessment. All three screens, regardless of our scoring methods, were significantly associated with the BDD, with explained variance ranging from 5% to 18%. MATILDA-R was associated with the highest percentage of explained variance (10% to 18%), perhaps due to similarities in task structure between the BDD and MATILDA-R. All three screens were significantly associated with the PSW (except for the WSD scoring option), but the WTW 18 Screen demonstrated low explained variance, making it impossible to establish a cut-off score to determine LD status. This was also true for the Empire State Screen when using the ETS scoring option. When the PSW determined LD status, the MATILDA-R demonstrated the highest percentage of explained variance (5% to 11%). The screens were also significantly associated with the DSM-5 approach, except for the MTE scoring option. The ETS scoring option on the Empire State Screen and the MTE scoring option on the MATILDA-R had explained variances too low to establish cut-off scores to determine LD. The WTW 18 Screen had the highest explained variance (6% to 12%) with DSM-5 determination.

Regarding each screen's sensitivity rating, the Empire State Screen lacked adequate sensitivity to correctly identify individuals who might be LD regardless of the LD determinant. Even though the Empire State Screen did attain a relatively reasonable overall

accuracy rate, it was not a viable option for identifying low-income Spanish speakers with LD as it did not appear to have adequate precision, at least given the set of individuals we tested.

Overall, the WTW 18 Screen also demonstrated low sensitivity when the LD status determinants were the BDD and the PSW diagnosis. However, when the LD status determinant was the DSM-5 diagnosis, the WTW 18 Screen demonstrated the highest level of sensitivity compared to the Empire State Screen and the MATILDA-R. Thus, we would recommend using the WTW 18 Screen with the WUS method of scoring (i.e., unweighted total score with a cut-off of 3 or more points) if the DSM-5 determinant is considered of primary importance.

Of the three screens, the MATILDA-R showed a more consistent pattern of accuracy and sensitivity across the three LD assessment methods, particularly when the LD status determinant is the PSW diagnosis. Thus, the MATILDA-R would provide a relatively moderate level of accuracy and sensitivity across the three LD status determinants depending on the scoring method employed. When we examine the PSW as the criterion, the MATILDA-R was the only measure to provide a reasonable level of sensitivity. Thus, based on our findings, MATILDA-R would be the best option if a clinical interview approach is preferred for diagnosing LD.

Screen Choice Considerations

On a practical level, the choice of a screening tool should take several factors into consideration. One is the uniqueness of the target population. Given differences in educational systems among various countries, it is unclear how strongly the U.S. educational standards and outcomes correspond to those of other nations, including those represented in the current sample. For example, the DSM-5 requires, for diagnosis of learning disorder, that the individual has a history of childhood learning problems. It may be particularly problematic to use this approach with the current population, as most of the participants completed their schooling in Latin America and had either a limited educational history (averaging 8th grade) or were never formally educated. Memories of childhood learning problems from early years may be particularly limited. Moreover, the current sample characterizes a unique segment of the Spanish-speaking population in the U.S. To an extent, research with the study sample reflects uncharted waters with respect to LD.

Because of the scarcity of studies available involving participants with characteristics like those in this sample, there is little other research to guide the selection of an LD screen for this population. Still, the specific characteristics of our sample should be kept in mind. Another factor is the method of LD diagnosis that clinical LD specialists use to determine LD status, which may vary depending on training and licensing requirements. Because the performance of each screen is dependent upon the LD diagnostic method, gaining information about commonly used clinical approaches will be important for making an informed decision about screen choices, including options for future research studies.

Moreover, the three screens were selected in part to be easy to administer, score, and interpret, which can be important considerations. Little training is required to administer, score, and interpret the WTW 18 Screen, with a cut-off score of 3 "yes" responses (unweighted). In contrast, the MATILDA-R requires more time to administer and score than does the WTW 18 Screen. The scored portion of the MATILDA-R consists of 8 "yes/no" background questions and 7 "yes/no" math-related questions. The MATILDA-R involves administration of 8 tasks that the individual is asked to complete. On average, the process takes 5 minutes. To administer and score the MATILDA-R, 3 to 4 hours of training and practice are required.

Implications for Social Work

Our findings are of interest to social workers (e.g., those in the field and those in policy positions) concerned with the Americans with Disabilities Act not being restricted from full implementation by assessments based on English language assumptions. As the U.S. has considerable language diversity (National Academies, 2017), it is likely that the full range of U.S. adults (and children) have not been screened accurately for possible learning disabilities, as many do not speak English well enough to demonstrate their disabilities on current screens and assessments (e.g., Chernoff et al., 2021). One important issue of relevance to social workers is the development of LD screens for Spanish-speaking adults who have the legal right to be appropriately assessed, as benefits (e.g., welfare) can be affected. Without appropriate screening and assessment, Spanish-speaking adults with LD cannot fully benefit from the protections afforded legally, nor the employment training and resources available to people with disabilities.

Limitations and Conclusion

Developing a brief screen to accurately predict LD status specifically (e.g., vs. "learning needs" or intellectual disability) is difficult. The overall accuracy rates of the screens here were generally 65% to 70%. Modifying the wording or scoring of test items may improve the precision of the screens. Moreover, the clinical specialist in our study did not interview participants. Clinical diagnosis of LD, traditionally, is a complex, expensive process that requires in-person testing and interviewing by a highly trained LD specialist. Thus, individuals identified as LD herein may not strictly meet federal guidelines for defining LD. We recognize, also, that linguistic equivalence through translation is often insufficient to guard against cultural bias and validity threats. There is a need to establish functional equivalence, cultural equivalence, and metric equivalence when assessments are translated to ensure validity and reliability of an LD screen. We do not assume that our translated measures are optimal indicators of LD. Instead, we attempted to examine their predictive validity (among this sample) regarding an LD diagnosis. We empirically tested whether these translated versions are appropriate for use as a screening measure for LD.

We limited the target population by using criteria for CalWORKs enrollment. Minimum snowball sampling (not strict probability sampling) was necessary to recruit eligible participants. Inferential statistics with data from a non-representative sample are fairly common in social science research (Rothman et al., 2013; TenHouten, 1992;

Trafimow, 2019). It can be questioned if inferences can be made beyond this sample. Finally, RAs were highly trained, but inter-rater reliability was not assessed. Future research should determine the findings' generalizability. Nevertheless, this study provides an initial estimate of the prevalence of LD risk in this population ($M = 21.9\%$; range $= 12\%$ to 36% depending on screen and cut-off method).

To conclude, the translated WTW 18 Screen and the MATILDA-R appeared most promising. A culturally-sensitive, validated LD screen will help ensure appropriate and legally required intervention and facilitate research for this understudied and underserved population.

References

Abwender, D. A. (2005). *Evaluation of screening devices for identifying learning disabilities in low-income Spanish-speaking adults*. SUNY at Brockport.

Agrawal, J., Barrio, B. L., Kressler, B., Hsiao, Y. J., & Shankland, R. K. (2019). International policies, identification, and services for students with learning disabilities: An exploration across 10 countries. *Learning Disabilities, 17*(1), 95-113. https://files.eric.ed.gov/fulltext/EJ1218057.pdf

American Psychiatric Association [APA]. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Author.

APA. (2013). *Diagnostic statistical manual: DSM-5*. Author.

Bronson, J., Maruschak, L. M., & Berzofsky, M. (2015). *Disabilities among prison and jail inmates, 2011–12*. US Department of Justice Bureau of Justice Statistics. https://bjs.ojp.gov/content/pub/pdf/dpji1112.pdf

Bujang, M. A., & Adnan, T. H. (2016). Requirements for minimum sample size for sensitivity and specificity analysis. *Journal of Clinical & Diagnostic Research, 10*(10), 1-6. https://doi.org/10.7860%2FJCDR%2F2016%2F18129.8744

California AB-1728 CalWORKs 2022 Reg. Sess. (CA. 2022). https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202120220AB1728

California Department of Social Services (2019, March). *CalWORKs California families working together: Annual summary.* https://www.cdss.ca.gov/portals/9/dssdb/calworksannualsummarymarch2019.pdf

Cameto, R., Knokey, A. M., & Sanford, C. (2011). Participation in postsecondary education of young adults with learning disabilities: Findings from NLTS2. *Learning Disabilities, 17*(2), 45-54. https://ies.ed.gov/ncser/pubs/20113005/pdf/20113005.pdf

Chernoff, J. J., Keuter, S., Uchikoshi, Y., Quick, H., & Manship, K. (2021). *Challenges in addressing California's diverse dual language learners*. American Institute for Research. https://californiadllstudy.org/sites/default/files/2021-02/Challenges%20in%20Assessing%20California%E2%80%99s%20Diverse%20DLLs.pdf

Corley, M. A., & Taymans, J. M. (2002). Adults with learning disabilities: A review of the literature. In J. Comings, B. Garner, & C. Smith (Eds.), *Annual review of adult learning and literacy* (vol. *3*, pp. 44-83). National Center for the Study of Adult Learning and Literacy. https://www.ncsall.net/index.html@id=575.html

Cortiella, C. (2013). Diplomas at risk: A critical look at the graduation rate of students with disabilities. National Center for LD. https://ncld.org/research/diplomas-at-risk-2/

Cortiella, C., & Horowitz, S. H. (2014). The state of learning disabilities: Facts, trends, and emerging issues. *National Center for Learning Disabilities, 25*, 2-45. https://www.ncld.org/wp-content/uploads/2014/11/2014-State-of-LD.pdf

Dondorf-Brooks, K., Seruya, F., Trenz, R., & Stanton-Koko, M. (2020). Parenting with a disability. *Journal of Applied Rehabilitation Counseling, 51*, 134-145. https://doi.org/10.1891/JARC-D-19-00014

Dorsett, J. (2000). A psychometric analysis of the Mississippi Assessment Technique for Identifying Learning Disabled Adults (MATILDA). *Dissertation Archive*. https://www.proquest.com/openview/01e106b9990d85552fa046f32754b86e/1?pq-origsite=gscholar&cbl=18750&diss=y

Department of Social and Health Services [DSHS]. (1998). *Learning disabilities*. Washington State Department of Social and Health Services, Economic Services Administration, WorkFirst Division.

Dunstan, D. A., & Scott, N. (2019). Clarification of the cut-off score for Zung's self-rating depression scale. *BMC Psychiatry, 19*(1), 1-7. https://doi.org/10.1186/s12888-019-2161-0

Dunstan, D. A., & Scott, N. (2020). Norms for Zung's self-rating anxiety scale. *BMC Psychiatry, 20*(1), 1-8. https://doi.org/10.1186/s12888-019-2427-6

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics, 21*(2), 553-565. https://doi.org/10.1093/bib/bbz016

Fiorello, C., Flanagan, D., & Hale, J. (2014). The utility of the pattern of the strengths and weaknesses approach. *Learning Disabilities, 20*, 55-59. https://doi.org/10.18666/LDMJ-2014-V20-I1-5154

Fletcher, J. M., Lyon, G., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities*. Guilford.

Fletcher, J., & Miciak, J. (2017). Comprehensive cognitive assessments are not necessary for the identification/treatment of learning disabilities. *Clinical Neuropsychology, 32*(1), 2-7. https://doi.org/10.1093/arclin/acw103

Fletcher, J., & Miciak, J. (2019). *The identification of specific learning disabilities: A summary of research on best practices*. Meadows Center for Preventing Educational Risk. https://files.eric.ed.gov/fulltext/ED606380.pdf

Franz, C., Ascherman, L., & Shaftel, J. (2017). *A clinician's guide to learning*

disabilities. Oxford University Press.
https://doi.org/10.1093/med/9780195383997.001.0001

Gerber, P. J. (2012). The impact of learning disabilities on adulthood: A review of evidence. *Journal of Learning Disabilities, 45*(1), 31-46.
https://doi.org/10.1177/0022219411426858

Goldberg, H. (2002). *Improving TANF program outcomes for families with barriers to employment*. Center on Budget & Policy Priorities.
https://www.cbpp.org/sites/default/files/atoms/files/1-22-02tanf3.pdf

Gregg, N., Coleman, C., Davis, M., Lindstrom, W., & Hartwig, J. (2006). Critical issues for the diagnosis of learning disabilities in adult population. *Psychology in Schools, 43*, 889-899. https://doi.org/10.1002/pits.20196

Gregg, N., Scott, S., McPeek, D., & Ferri, B. (1999). Definitions and eligibility criteria applied to the adolescent and adult population with learning disabilities across agencies. *Learning Disability Quarterly*, *22*, 213-223.
https://doi.org/10.2307/1511287

Grigorenko, E. L., Compton, D. L., Fuchs, L. S., Wagner, R. K., Willcutt, E. G., & Fletcher, J. M. (2020). Understanding, educating, and supporting children with specific learning disabilities: 50 years of science and practice. *American Psychologist, 75*(1), 37-51. doi.org/10.1037/amp0000452

Grubb, R., Hemby, V., Walker, J., & Pierce, W. (1997). MATILDA: The development of a learning disabilities screening device. *Adult Basic Education, 7*(1), 23-38.
https://www.proquest.com/openview/6a7a1ea2abf967b2536e28dffcbe4200/1?pq-origsite=gscholar&cbl=30221

Grubb, R., Hemby, V., Walker, J., & Pierce, W. (2001). *MATILDA*. University of Southern Mississippi.

Holcomb, P., & Thompson, T. (2000). *State welfare-to-work policies for people with disabilities.* U.S. DHHS, Office of the Assistant Secretary for Planning and Evaluation. https://www.urban.org/sites/default/files/publication/62016/1000043-State-Welfare-to-Work-Policies-for-People-with-Disabilities.PDF

Holliday**,** G., Koller, J., & Thomas, C. (1999). Post–high school outcomes of high IQ adults with LD. *Journal of the Education of Gifted, 22*, 266-281.
https://doi.org/10.1177/016235329902200303

Horowitz, S. H., Rawe, J., & Whittaker, M. C. (2017). *The state of learning disabilities: Understanding the 1 in 5*. National Center for Learning Disabilities.
https://www.ncld.org/news/newsroom/the-state-of-ld-understanding-the-1-in-5/

Individuals with Disabilities Education Improvement Act (IDEA), Public Law 108-446, U.S. Code 20, § 1414 (2004). https://www.congress.gov/108/plaws/publ446/PLAW-108publ446.pdf

Jiménez, J., & García Cadena, C. (2007). Learning disabilities in Guatemala and Spain.

Learning Disabilities Research & Practice, 22, 161–169.
https://doi.org/10.1111/j.1540-5826.2007.00247.x

Johnson, T. P. (2014). Snowball sampling: Introduction. *Wiley StatsRef: Statistics Online*. https://doi.org/10.1002/9781118445112.stat05720

Johnson, E., Humphrey, M., Mellard, D., Woods, K., & Swanson, H. (2010). Cognitive processing deficits and students with specific learning disabilities. *Learning Disability Quarterly, 33*, 3-18. https://doi.org/10.1177/073194871003300101

Kauf, J. (2008). *Assisting TANF recipients living with disabilities to obtain and maintain employment: Final report*. https://www.mathematica.org/download-media?MediaItemId={4E2379E2-5A42-41A1-AF1F-774D836ED21F}

Klassen, R. M., Tze, V. M. C., & Hannok, W. (2011). Internalizing problems of adults with learning disabilities: A meta-analysis. *Journal of Learning Disabilities, 46*, 317-327. https://doi.org/10.1177/0022219411422260

Kusserow, R. P. (1992). *Functional impairments of AFDC clients*. US DHHS, Office of the Inspector General. https://oig.hhs.gov/oei/reports/oei-02-90-00400.pdf

Lancaster, S. J. (2002). *Determining validity and reliability of an online learning disabilities screening battery for use on adults* [Dissertation]. University of Kansas. https://www.proquest.com/openview/1efa0bd03dbaa53e6c6782dfd6f472ed/1?pq-origsite=gscholar&cbl=18750&diss=y

Lightfoot, A., Janemi, R., & Rudman, D. L. (2018). Perspectives of North American postsecondary students with learning disabilities: A scoping review. *Journal of Postsecondary Education and Disability, 31*(1), 57-74. https://files.eric.ed.gov/fulltext/EJ1182368.pdf

Loprest, P., & Maag, E. (2001). *Barriers to and supports for work among adults with disabilities*. The Urban Institute. https://www.urban.org/sites/default/files/publication/61576/410107-Barriers-to-and-Supports-for-Work-among-Adults-with-Disabilities.PDF

Maki, K. E., Floyd, R. G., & Roberson, T. (2015). State learning disability eligibility criteria: A review. *School Psychology Quarterly, 30*(4), 457-469. https://doi.org/10.1037/spq0000109

Miciak, J., & Fletcher, J. (2019). The identification of learning disabilities. In M. Burns (Ed.), *Introduction to school psychology* (pp. 336–359). Oxford University Press.

Miciak, J., & Fletcher, J. M. (2020). The critical role of instructional response for identifying dyslexia and other learning disabilities. *Journal of Learning Disabilities, 53*(5), 343-353. https://doi.org/10.1177/0022219420906801

Miles T. R. (2004). Some problems in determining the prevalence of dyslexia. *Electronic Journal of Research in Educational Psychology*, *2*, 5-12. https://eric.ed.gov/?id=EJ802296

Moats, L. C., & Dakin, K. E. (2008). *Basic facts about dyslexia and other reading*

problems. International Dyslexia Association.

Muñoz-Sandoval, A. F., Woodcok, R. W., McGrew, K. S., & Mather, N. (2005). *Batería III Woodcock-Muñoz Pruebas de aprovechamiento*. Riverside.

National Academies of Sciences, Engineering, and Medicine. (2017). *Promoting the educational success of children and youth learning English*. National Academies Press.

Newman, L., Wagner, M., Cameto, R., Knokey, A., & Shaver, D. (2010). *Comparisons across time of the outcomes of youth with disabilities up to 4 years after high school*. SRI International. https://files.eric.ed.gov/fulltext/ED512149.pdf

Pavetti, D. (2018). *TANF studies show work requirement proposals for other programs would harm millions, do little to increase work*. Center on Budget and Policy Priorities. https://www.cbpp.org/sites/default/files/atoms/files/11-13-18tanf.pdf

Rothman, K. J., Gallacher, J., & Hatch, E. E. (2013). Why representativeness should be avoided. *International Journal of Epidemiology, 42*, 1012-1014. https://doi.org/10.1093/ije/dys223

Seo, Y., Abbott, R. D., & Hawkins, J. D. (2008). Outcome status of students with learning disabilities at ages 21 and 24. *Journal of Learning Disabilities, 41*, 300-314. https://doi.org/10.1177/0022219407311308

Shapiro, J., & Rich, R. (1999). *Facing learning disabilities in the adult years. Understanding dyslexia, ADHD, assessment, intervention, and research*. Oxford University Press.

Shifrer, D., Muller, C., & Callahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, SES, and language. *Journal of Learning Disabilities, 44*(3), 246-257. https://doi.org/10.1177/0022219410374236

Sideridis, G. D. (2007). International approaches to learning disabilities: More alike or more different. *Learning Disabilities Research & Practice, 22*, 210–215. https://doi.org/10.1111/j.1540-5826.2007.00249.x

Skellern, J., & Astbury, G. (2014). Gaining employment: The experience of students at college with LD. *British Journal of Learning Disabilities, 42,* 58-65. https://doi.org/10.1111/bld.12012

Stough, L., & Aguirre-Roy, A. (1997). Learning disabilities in Costa Rica. *Journal of Learning Disabilities, 30*(5), 566-571. https://doi.org/10.1177/002221949703000513

Swanson, H. L., Harris, K., & Graham, S. (Eds.). (2013). *Handbook of learning disabilities*. Guilford.

Sweeney, E. P. (2000). *Recent studies indicate that many parents who are current or former welfare recipients have disabilities or other medical conditions*. Center on Budget and Policy Priorities. https://www.cbpp.org/2-29-00wel.pdf

Taymans, J. M. (2009). *Learning to achieve*. National Institute of Literacy.

http://lincs.ed.gov/publications/pdf/L2ALiteratureReview09.pdf

Taymans, J. (2012). Legal and definitional issues affecting the identification and education of adults with specific learning disabilities. *Journal of Learning Disabilities*, *45*, 5-16. https://doi.org/10.1177/0022219411426857

TenHouten, W. (1992). Generalization and statistical inference from snowball samples. *Bulletin of Sociological Methodology, 37*, 25-40. https://doi.org/10.1177/075910639203700102

Trafimow, D. (2019). Five nonobvious changes in editorial practice for editors and reviewers to consider when evaluating submissions in a post *p* < .05 universe. *The American Statistician, 73*, 340-345. https://doi.org/10.1080/00031305.2018.1537888

U.S. Census Bureau. (2021). *Racial and ethnic diversity in the United States: 2010 Census and 2020 Census.* https://www.census.gov/library/visualizations/interactive/racial-and-ethnic-diversity-in-the-united-states-2010-and-2020-census.html

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Rehabilitation Services Administration. (2022). Section 300.307 Specific learning disabilities. https://sites.ed.gov/idea/regs/b/d/300.307

U.S. DHHS, Office of Family Assistance. (2021). *Characteristics and financial circumstances of TANF recipients, fiscal year 2020.* https://www.acf.hhs.gov/ofa/data/characteristics-and-financial-circumstances-tanf-recipients-fiscal-year-2020

Wagner, M., Newman, L., Cameto, R., Garza, N., & Levine, P. (2005). *After high school: A first look at the postschool experiences of youth with disabilities*. SRI International. http://files.eric.ed.gov/fulltext/ED494935.pdf

Williams, J. L., Miciak, J., McFarland, L., & Wexler, J. (2016). Learning disability identification criteria and reporting in empirical research: A review of 2001–2013. *Learning Disabilities Research & Practice, 31*(4), 221-229. https://doi.org/10.1111/ldrp.12119

Young, G., & Browning, J. (2005). Learning disabilities/dyslexia and employment. In G. Reid & A. Fawcett (Eds.), *Dyslexia in context* (pp. 25-59). Whurr Publishers.

Zimmerman, M., & Galione, J. N. (2011). Screening for bipolar disorder with the Mood Disorders Questionnaire: A review. *Harvard Review of Psychiatry, 19*(5), 219-228. https://doi.org/10.3109/10673229.2011.614101

Author note: Address correspondence to Gail S. Goodman, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616. Email: ggoodman@ucdavis.edu