DEVELOPMENTAL DIFFERENCES IN DISABILITY EVALUATIONS WITH ENGLISH LEARNERS: ARE PSYCHOLOGISTS CONSIDERING THEM?

A dissertation submitted in partial fulfillment
of the requirements for the degree of

DOCTOR OF PSYCHOLOGY

to the faculty of the

DEPARTMENT OF PSYCHOLOGY

of

ST. JOHN'S COLLEGE OF LIBERAL ARTS AND SCIENCES

at

ST. JOHN'S UNIVERSITY
New York
by
Erick Bienvenido Díaz

Date Submitted 5/24/2024 Date Approved 8/11/2024

Erick Bienvenido Díaz	Samuel O. Ortiz, Ph.D.

ABSTRACT

DEVELOPMENTAL DIFFERENCES IN DISABILITY EVALUATIONS WITH ENGLISH LEARNERS: ARE PSYCHOLOGISTS CONSIDERING THEM?

Erick Bienvenido Díaz

The main purpose of this pilot study was to explore the rate and extent to which school psychologists are considering different developmental, linguistic, and environmental factors of English Learners. The sample of this study included coding bilingual psychoeducational evaluations (N=120) for such factors using the D-STPGE scale as a guide to explore the consideration of each factor. The pool of reports included bilingual psychoeducational, educational, and neuropsychological evaluation for bilingual (Spanish) children between kindergarten and 6th grade. The procedures of this research include understanding the frequency of each reported construct paired with a chi-square legacy test to explore if the observed construct is being reported within the expected statistical frequency or outside of such realm. Such constructs that were more likely to be considered include heritage/home language, length of learning English, and current English language development. Constructs that evaluators had poor compliance in reporting include the examinee's home culture match to the US middle class, socioeconomic status, community language, and support for social-emotional development. Finally, constructs that were within the expected statistical frequency include school attendance/participation, atypical developmental life experiences, and

formal education in the heritage language. Further research is needed to understand the psychometric properties of the D-STPGE. However, this study helps to understand the gaps in training and report writing within nondiscriminatory and bilingual evaluations. Overall, school psychologists who conduct bilingual and multilingual/nondiscriminatory assessments need a guide like the D-STPGE to be able to compare and quantify the difference that exists between and amongst others within their respective peer groups.

ACKNOWLEDGEMENTS

I would like to express my deepest appreciation to the SJU faculty and program, including having the opportunity to work closely with Dr. Samuel Ortiz as my dissertation mentor along with my dissertation committee members, Dr. Flanagan Orgiz and Dr. Marlene Sotelo-Dynega. I could not have undertaken this journey without the help and guidance of my dissertation mentor. Furthermore, I would like to express gratitude to all of my professors, supervisors, mentors, classmates, and colleagues who have helped shape me into the clinician that I am today. Furthermore, I would like to acknowledge my parents Betty and José Joaquín, my stepfather, and my siblings, Eduardo, Madelyn, Daniela, and José, for their moral support and encouragement for my endeavors in pursuing a doctoral degree. I also dedicate this dissertation to my nieces and nephews, who have inspired me to pursue a degree in Psychology where I can be able to utilize my clinical skills to help English Learners reach their full potential. Last but not least, I'd like to give many thanks to my close friends who also kept me grounded during this long, challenging, but overall rewarding process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Recently published survey data from 2000-2017 by the U.S. Department of Education demonstrates that 5 million students in the U.S. K-12 education system are English Learners (or ELs) (2020). Despite being the most rapidly growing student population in elementary and middle schools, many of these students are at risk for achievement problems and dropping out of school (Sheng et al., 2011). The complex nature of education for a culturally and linguistically diverse child has been a long-standing issue in public schools. Many of these issues arise when it comes to deciphering the child's learning needs especially when it comes to language development. Despite the growing percentage of ELs in the U.S., the educational system still has difficulty assessing linguistically diverse children who present with learning difficulties (Garcia et al., 2008, p. 9; Sullivan, 2011). This is often cited as one of the main reasons why this demographic is either over or underrepresented when it comes to special education placement and services.

A plethora of intervening techniques exist to evaluate children's cognitive and achievement ability, including standardized, norm-referenced cognitive testing, achievement testing, and language testing. However, common psychological batteries that are used to assess learning issues in schools have sample norms where many individuals are monolingual and monocultural (Norfolk et al, 2015; Garcia et al, 2008, p. 16). With but two exceptions, even the inclusion of English learners in a sample is little more than window dressing as there is no differentiation between them in terms of experience and development in the language of the test, despite being of the same age (Ortiz, 2019; Ortiz & Cehelyk, 2023). Furthermore, there are no standardized procedures

regarding the evaluation of a child who is learning English. Many districts handle these issues on a case-by-case basis and determine for themselves what to do in idiosyncratic ways (Garcia et al., 2008).

Regardless of the manner or methods employed in the evaluation of ELs, particularly for disability identification and special education eligibility determination, there continues to be a lack of attention paid to the developmental and experiential differences that characterize and differentiate English learners from monolinguals (Fisher & Frey, 2012; Ortiz & Cehelyk, 2023; Salvia & Ysseldyke, 1991; Wolfram, Adger & Christian, 1999). In conjunction with standard practices for nondiscriminatory and bilingual assessments, it is important to include information regarding the individual's generational history, familial history, language proficiency and acquisition, socio-economic status, opportunity to learn, academic history, as well as an understanding of the individual's experiences during developmental periods of life (Rhodes, Ochoa, & Ortiz, 2005). Such differences involving language and acculturative knowledge acquisition represent circumstantial factors that determine the extent of differences that may exist even among individuals of the same age (Ortiz, 2019). Furthermore, the psychological tools employed in a psychoeducational evaluation possess items and procedures that are consistent with the culture from which the norms have been collected; the less exposure an individual has had to a different culture, the larger the "difference" can be noted within specific constructs and the individual's performance on cognitive measures (Vazquez-Nuttall et al, 2007, p. 266). Without a careful and detailed examination of the extent to which such factors have been present in the life of an EL, there is no possibility that any evaluation of the child will be valid or useful for

identifying whether an actual disability exists or whether the manifest learning problems are more the result of some type of difference. Therefore, the purpose of this study is to closely examine the extent to which school psychologists who are conducting this type of evaluation are providing sufficient attention to developmental differences that have a direct bearing on the validity of any obtained test results as well as any subsequent interpretation drawn from them.

LITERATURE REVIEW

Concerning issues with suspected disability and differences in learning experiences, research has long highlighted an overidentification of minorities in special education. Many factors exist as to why the number of minority students in special education is not proportionate to the general population: language abilities, bilingualism, socioeconomic status, poor sociocultural interpretation of psychological testing, comparing a student's performance to norms that do not reflect the specific child's experience, and so forth. Dating back to the beginning of IQ testing, it has been documented that many immigrants and African Americans were deemed "mentally incompetent" since their performances on early measures of intelligence tests did not reflect what was deemed intelligent (i.e., their own culture, language, etc.; Reddy, 2008) Many studies examining the differences in language and different dialects of English have debunked the notion that dialect speakers are somehow deficient: "The key consideration in distinguishing between a difference and a disorder is whether the child's performance differs significantly from peers with similar experiences." (Wolfram, Adger & Christian, 1999, p.105). The topics of individual and developmental differences have shed light on the notion that research, practice, and policy should adhere less to standard "categorizations" and more "inclusive and cognizant" of individualistic features of childhood development and early learning skills (Institute of Medicine and National Research Council, 2015, pg. 87). This includes the rising belief that such differences should be considered mutually inclusive and part of a child's cultural development instead of hierarchical and deficient because it is different from the general population (i.e., African American Vernacular English vs. Standard American English).

Additionally, issues like misinterpretations of behaviors, insufficient professional training on diverse populations, poverty, as well as lack of formal instruction are factors that have been continuously deemed the culprits as to why many schools tend to have high referral rates for minority students (Kreskow, 2013). This is especially true for ELs, who might appear inattentive in a classroom, may have difficulties following directions, may have grown up in a community with a lack of resources and programs, may lack formal education in their native language, and may also score low on verbal measures (due to the lack of knowledge and acculturative experiences in English) and will possibly get referred for suspected disability despite these issues being a difference in cultural and language. These factors are not intrinsically an issue of racial differences, but rather, layers of acculturative knowledge and differences in development that shape and mold a child's presentation within the social and educational environment in the U.S. In short, the issue is whether these individualistic factors are being considered when diverse individuals are evaluated for special education services. Overall, evaluators conducting assessments on ELs suspected of having a disability should consider these different elements to defend their findings as non-discriminatory and valid. However, there is a lack of research that explores to what extent are these factors reported and considered and which factors tend to be more considered.

Over and Underrepresentation of ELs in Special Education

It is a commonly known issue that many minorities are overrepresented and underrepresented in special education. Many of the minorities included in overrepresentation happen to be individuals who are learning English: many schools and

districts have differing ways in which they assess ELs and often struggle to find the most appropriate instructional program (Burr, Hass, & Ferriere, 2015). Due to decades of poor evaluation techniques, lack of funds, and inherited "systemic bias", there has been an overrepresentation of English Language Learners in special education which has been deemed "discriminatory" (Skiba, Poloni-Staudinger, Simmons, Feggins-Azziz & Chung, 2005, p. 142). However, there exist many mechanisms that have analyzed culturally and linguistically diverse children. Research has demonstrated that in many areas of the U.S.A., minority children are underrepresented and less likely to be referred for services or evaluations (Hibel et al., 2010; Delgado & Scott, 2006). The difficulty of this issue arises from the complex fact that academic difficulties are linked to learning disabilities, emerging language acquisition, cultural differences in academic instruction, or even a combination of all these issues (Shore & Sabatini, 2009). Many reasons as to why there is an over and underrepresentation of EL in special education are due to bias (i.e., cultural bias, racial bias, lack of judgment regarding the development of bilingual children, etc.), improper testing, and statewide policies (i.e., only using results from standardized testing measures to diagnose disability), as well as societal views on children who are culturally and linguistically diverse (Harry & Klingner, 2006). Many studies have demonstrated issues with the underrepresentation of ELs at the elementary and secondary level due to poor evaluation, testing measures, and exiting students prematurely: This can indicate that many ELs require special education and extra support but are not receiving it due to this issue (López & Linn, 2018). Furthermore, ELs tend to be underrepresented on special education rosters (since their educational plan possibly focuses on ESL services) but tend to be overrepresented in different special education categories like Learning Disability,

Intellectual Disability, and Emotional disturbance (DeMatthews et al, 2014). Less representation is seen in higher grades of education, and many believe it might also be due to a lack of a pre-referral process on top of identification and assessment issues (DeMatthews et al, 2014). Language minorities possess basic interpersonal communication skills (BICS) that can be used to help them acquire knowledge of English: Many educators fail to understand and evaluate the different degrees of literacy and oral fluency that many ELs possess (August & Shanahan, 2006). This information can be used to tailor interventions and create advantages for English education. This lack of understanding along with linguistic and immigration factors can lead to many children being identified for special education when it is not warranted (Rueda and Windmuller, 2006). Proficiency in the first language also has developmental implications for the student's acquisition of their second language; otherwise known as Cognitive Academic Language Proficiency or CALP (Cognitive Academic Language Proficiency; August & Shanahan, 2006). For example, much research has demonstrated students have higher academic success in learning English when they possess cognitive academic language proficiency skills in their native language (August & Shanahan, 2006). To improve the issue of overrepresentation and underrepresentation of EL in Special Education Services, the education system must improve the ways they deliver culturally sensitive evaluation methods that can be tailored to the student's educational needs.

Despite much research on ELs and educational accountability in the United States of America for every student, there is no standard definition to identify them or interventions that have been deemed ideal for this type of minority student (Clark-Gareca et al., 2019). This may result in many states taking different approaches to evaluate and

assign services to an EL with no precise protocol, guidance, or standard. This can impact the student's long-term academic achievement, especially if the student is not making progress because of certain interventions. It also becomes a greater issue if the extra academic support is not delivered in their native language or learning style. Although it is widely known that it is ideal to test the student in their native language, this is usually not feasible due to the lack of practitioners that may speak the student's native language and the availability of standardized tests that have been normed in the student's native language (Chu & Flores, 2011). Technology like the Cross Battery Assessment Software System (X-BASS; Flanagan, Ortiz, Alfonso, 2013) can help school psychologists distinguish between ability and knowledge that has not been hindered by cultural and linguistic differences in a child with the use of the Culture-Language Interpretive Matrix (C-LIM; Flanagan, Ortiz & Alfonso, 2013; Ortiz, 2019) (Cormier, McGrew and Ysseldyke, 2014, Cormier et al., 2022). The use of X-BASS and the C-LIM helps practitioners with a comprehensive framework when it comes to assessing culturally and linguistically diverse children. Furthermore, it helps make a distinction between difference versus disorder by considering a student's linguistic background and comparing them to other culturally and linguistically diverse students' performance on cognitive measures that load heavily on language and culture compared to those that are more nonverbal in nature (Flanagan, Ortiz, & Alfonso, 2013). However, an evaluation should always consider differences in the native language, educational experience in all spoken languages, age, and overall development when assessing an English language learner (Bernan, Haertel, & Pellegrino, 2020, p. 152). This is not the case when it comes to the administration of the NYSITELL and NYSESLAT. This brings about the issue

with assessing English language learners and how most assessments do a poor job of determining the child's strengths and weaknesses of their learning capacity or how they will perform over time on reading and content assessment in English (August & Shanahan, 2006). As the population of ELs in the U.S.A. grows, schools and districts must be doing their due diligence to provide adequate services to these individuals. This includes the need for practitioners to familiarize themselves with standardized testing and interpreting results via a nondiscriminatory lens that considers the child's cultural and linguistic background.

Evaluating English Learners and Test Performance

The issue with testing ELs is that many standardized measurements have content that is attributed to broad experiential factors, primarily cultural and linguistic differences. The question that comes to mind is if the examinee has a learning disability or has not acquired enough English to perform well during the evaluation of broad cognitive abilities. In this instance, validity is what's most important: when it comes to measuring a construct that we believe we are measuring (i.e., verbal comprehension, reasoning abilities, working memory, etc.), many psychological batteries are structured in a psychometric way that does not make it possible for performances in the average range and above to be due to chance (Ortiz et al, 2016). Construct validity refers to whether a scale or psychological test measures the construct adequately. Although tests are not inherently biased, they usually demonstrate scores that do not paint an accurate picture of a culturally or linguistically diverse child: Since many tests have been normed on the public, performances by minorities tend to be lower than expected, demonstrating

cultural and linguistic bias (Jensen, 1976; Espinoza, 2010; Sotelo-Dynega et al, 2013). This can also mean that low test performance can be due to issues other than below-average ability (i.e., cultural, or linguistic factors). As expressed by Ortiz and his colleagues (2016), validity is primarily an issue when performance on subtests is below average, and follow-up is needed to differentiate between true ability and cultural or linguistic factors. For example, subtests that measure learning and retrieval fluency (G*l* *and* G*r*) rely heavily on the use of language and teaching the examinee task demands. Furthermore, a large portion of intelligence testing relies on our use of language, which naturally is culturally bound. This can create issues with ELs who have not had the same amount of English exposure compared to same-age and same-grade peers that standardized tests have been normed upon. Although strictly using nonverbal tests on ELL sounds like a great alternative to this current issue, some evidence exists that nonverbal tests yield low validity in predicting reading and math abilities (Lakin, 2012; Lochman et al., 2008). However, language-reduced cognitive measures are just as important as verbally loaded tests since they paint a picture of the linguistically diverse individual if scores are interpreted qualitatively and independently, rather than interpreting quantitative scores and simply deeming them valid for an individual that is not a part of the mainstream culture.

Experiential and Developmental Factors that Impact Learning Acquisition

Many of the current evaluation techniques for language minorities in the school system fall short for many English language learners who are not meeting academic expectations and need tailored interventions. Most standardized assessments are normed

using monolingual English speakers who have only had exposure to American education. Issues related to differences in the acculturation of learning experiences and language proficiency in the developmental stages of monolingual and monocultural English speakers are controlled by psychometrics and other statistical procedures that create norms for standard cognitive batteries: This cannot be said nor utilized in the same framework when evaluating individuals who are culturally and linguistically diverse (Ortiz, Ortiz, & Devine, 2016; Cormier, McGrew, & Ysseldyke, 2014). Furthermore, these standardized assessments do not consider the student's proficiency in the English language, prior education history, as well as the student's proficiency (both oral and academic) in their language (MacSwan & Rolstad, 2006; Chu & Flores, 2011). As expressed by Freeman & Freeman (2004), students who are ELs fall into four categories: a student newly arrived with some exposure to schooling, a student newly arrived with limited or no exposure to schooling, a student exposed to two languages at once, and a student who is a long-term English language learner. Recently, standardized language proficiency tests have been developed like the BESA and Ortiz PVAT that can provide estimates of language development for children who speak more than one language (Alfonso et al., 2020). This demonstrates a step forward towards fair assessment and controlling aspects of validity that have been overlooked since most psychological batteries have been normed on a monolingual English-speaking population. This is different from the NYSITELL and NYSESLAT tests, which state that the tests measure "English language proficiency" but do not specifically explain the construct of proficiency and how it fits in an academic context (Gareca, 2016, p. 37). Furthermore, Gareca highlights that inconsistency and lack of clarity on proficiency scores and added

pressure to meet a quota for exiting students out of ESL services can have the potential to impact the learning of EL students (2016). The NYSITELL and NYSESLAT can improve their testing measures by clearly defining the construct of language, increasing the reliability and validity of test norms, and possessing more rigid standardized test administrations (NYSUT, 2015; Gareca, 2016). What every school psychologist and evaluator should consider is analyzing multiple data points for evaluations to tailor specific needs for students who are learning English if testing norms do not have a representative sample of ELs. For example, because many English learners are from a different country and/or culture, their experiences at home in terms of what they hear, what they learn, what they are exposed to, and so forth, create differences in what they ultimately bring to school or the testing environment. The importance of such experiential factors was identified early on but not necessarily paid attention, as Sanchez (1934) argued, "As long as tests do not at least sample in equal degree a state of saturation [assimilation of fundamental experiences and activities] that is equal for the 'norm children' and the bilingual child it cannot be assumed that the test is a valid one for the child." (p. 770). In short, it is important to compare a child's individualistic and developmental factors to others who have had similar experiences to gauge whether that child's capacity to learn in the community is truly deficient or different. This same point was further reiterated by Salvia and Ysseldke (1991),

> When we test students using a standardized device and compare them to a set of norms to gain an index of their relative standing, we assume that the students we test are similar to those on whom the test was standardized; that is, we assume their acculturation is comparable, but not necessarily

> identical, to that of the students who made up the normative sample for the test. When a child's general background experiences differ from those of the children on whom a test was standardized, then the use of the norms of that test as an index for evaluating that child's current performance or for predicting future performances may be inappropriate. Incorrect educational decisions may well be made. (p. 18).

However, it is also important to recognize that experiential differences are not the same as racial differences or merely differences in language. Salvia and Ysseldyke (1991) further emphasized this point by highlighting the problems with the assumption of comparability in testing as they relate to development, not to socially constructed categories. They stated,

> It must be pointed out that acculturation is a matter of experiential background rather than of gender, skin color, race, or ethnic background. When we say that a child's acculturation differs from that of the group used as a norm, we are saying that the experiential background differs, not simply that the child is of different ethnic origin, for example, from the children on whom the test was standardized (p. 18).

In short, to accurately and validly measure an individual's abilities, knowledge, or skill, they should be compared to norms, populations, and other individuals who have had similar exposure to developmental experiences including similar opportunities for the acquisition of the language of the test and

the cultural knowledge that may be embedded in it Language is perhaps the single most important developmental factor in the testing process as it is used as an integral method for conducting an evaluation. One cannot be tested without communication of some sort, and even when that communication is designed to be simple, or there is a belief regarding sufficient proficiency, such assumptions may still not hold. Cormier and colleagues (2022) assert that "although a student's conversational level of English language proficiency could be perceived to be relatively consistent with their peers', their level of academic language proficiency may not be sufficient to fully benefit from classroom instruction or understand test directions to the same extent of a native English language speaker." English learners simply cannot be viewed as being comparable in their language development when compared to monolinguals who have spent their entire lives with only one language—that is, the language of the test. In contrast, English learners have considerable variability in how long they have been learning the language of the test and are not comparable even among themselves due to significant differences in when each began learning the language of the test. As Fisher and Frey (2012) pointed out, "It is unlikely that a second-grade English learner at the early intermediate phase of language development is going to have the same achievement profile as the native English-speaking classmate sitting next to her. The norms established to measure fluency, for instance, are not able to account for the language development differences between the two girls. A second analysis of the student's progress compared to linguistically similar students is warranted." (p. 40). When ELs progress is compared to the progress of

other students who have had longer exposure to English and acculturative knowledge, this interpretation of their performance is likely discriminatory. Language is a strong factor of intelligence and acculturative knowledge; almost all norms on standardized measures only control for individuals who have been exposed to one language throughout their whole development.

Other important factors that impact education and acculturative knowledge happen to be socioemotional functioning and atypical developmental experiences in childhood. Throughout research in the public school system within urban environments, there has been a great concern for students of color who not only live in poverty but have also been exposed to trauma. Trauma can include any violence within the home or community, exposure to deportation, having a status as a refugee, prior persecution in the country of origin, and so forth. Experiences with such events harm the student's well-being and very much lead to negative outcomes concerning their learning capacity, social-emotional functioning, and conduct within the school system (Blitz, Anderson, and Saastamoinen, 2016). Research has also demonstrated that those who live in poverty are more likely to experience adverse events. Concerning psychological research, it is important to consider these factors because if a child is presenting with many difficulties that are likely precipitated by trauma and lack of social-emotional support, they should not be classified as learning disabled. It is pertinent that psychologists are considering and documenting such sensitive information.

PRESENT STUDY

To understand the performances of ELs on measures that rely on developmental expectations of the English language and acculturative knowledge acquisition (e.g., cognitive abilities, academic skills, general knowledge, language), it is important to consider relevant factors that create the context by which psychologists can interpret their results fairly. Recent research has been highlighting the importance of understanding the examinee's characteristics (i.e., language development) over the characteristics of psychometric tests (Cormier et al., 2022). Individual and developmental factors captured on the Diverse Student True Peer Group Estimator (D-STPGE; Ortiz, 2022) include heritage/home culture, socioeconomic status, home language, school attendance, atypical developmental life experiences (i.e., refugee, exposure to war, trauma, etc.), formal education in the native language, length of exposure to English, current English language development, community language, and support for social-emotional development. Given the importance of these factors, as listed on the D-STPGE, and their bearing on the validity of any collected data and obtained test scores, it is expected that they should be reported and discussed in psychological reports in a manner that bolsters the credibility of any psychological interpretations and meaning that are offered in the assessment of culturally and linguistically diverse individuals. Conversely, any failure to note or explain whether and to what extent such developmental experiences have affected the validity of the obtained results and subsequent interpretations would constitute legally indefensible conclusions based on the premise that, if it is not documented, then it did not happen. Therefore, the present study is exploratory and seeks to determine whether psychologists are noting and discussing any of the ten elements of developmental differences as

contained in the D-STPGE. Evaluation of compliance in this regard can be measured by simple count and frequency as documented within the body of a report of psychoeducational evaluation. This study thus seeks to measure the extent to which psychologists are considering the various aspects of language and acculturative knowledge acquisition, as well as other developmental factors common to and often present in the assessment of diverse individuals. Ultimately, it can be reasonably assumed that the greater the number of issues identified and discussed within the context of an evaluation report, the greater the likelihood that the conclusions and meaning assigned to the data are valid. Should reports of evaluation fail to provide mention or discussion of these issues, it can be assumed that the evaluation lacks significant validity and any meaning assigned to the collected data would lack validity and be spurious, at best.

Research Question

It is expected that all psychologists have considered all potentially relevant individualistic and developmental factors that are described in the D-STPGE and have documented these factors and their impact in their evaluation reports as would be considered best practice. Examination of compliance with these criteria will, in turn, provide a frequency count for each factor described in the D-STPGE which would be resumed to be equal across all areas, given that no single one is considered substantially more important than any other. Despite this positive expectation, one or more factors are not being properly documented or considered. Essentially, the final frequency distribution would help explore the extent to which each factor is or is not being considered and reported within the context of the written reports of psychological evaluations on

culturally and linguistically diverse children. In short, if psychologists do not document the extent to which they have considered all of the relevant experiential factors that may affect the evaluation of an EL's abilities, there is no guarantee that the evaluation has been conducted in a fair and nondiscriminatory manner and might serve as an indictment of the current procedures and methods being used by school psychologists in their evaluations of English learners.

METHODS

Participants

This study utilized archival data which includes de-identified psychological, psychoeducational, and neuropsychological reports. The sample includes reports (N=120) collected from the state of California to analyze whether and to which extent were individual characteristics and developmental factors considered when evaluating culturally and linguistically diverse children and young adults.

Procedures

To evaluate the different characteristics of English learners that should be considered when evaluating culturally diverse children and young adults, the DSTPGE scale was utilized to analyze the archival data. The archival data consists of redacted reports from bilingual psychoeducational evaluations from the state of California conducted between 2022-2023. The D-STPGE possesses 10 individual items and a code of 1 (item or construct indicated in the psychoeducational report) or 0 (item or construct not indicated in the psychoeducational report) will be indicated on each item of the DSTPGE for each psychoeducational report analyzed.

Tools and Instruments

Diverse Student True Peer Group Estimator (D-STPGE). The D-STPGE is a scale that can be used to guide clinical decision-making when interpreting expectations of performance with culturally and linguistically diverse individuals. The “degree of difference” provided can be used to establish the amount of impact such variables might

have had on the collection of any type of data, including pre-referral measurements within an MTSS system, or even after administration of standardized tests. The degree of difference is, in fact, a requirement for the use of the C-LIM and C-LIM+ATE when evaluating the impact of cultural and linguistic factors on cognitive and academic test performance respectively.

The scale is comprised of ten questions that aim to quantify the student's true peer group. Each item on the D-STPGE focuses on different developmental factors that may well impact culturally and linguistically diverse students' achievement in academics. It is a way to compare how culturally, and linguistically diverse students are similar or different to the general population of students that they tend to be compared to. Questions included in this scale gauge the student's heritage, social economic class, exposure to English and heritage language instruction, unusual or atypical developmental life experiences, academic levels of English proficiency, and community values where the student resides, amongst other factors that may affect expectations of development, learning, and growth. This study aimed to quantify the frequency with which psychologists consider each facet concerning bilingual and nondiscriminatory psychoeducational and neuropsychological assessments. Since the nature of this study is a frequency distribution of the important developmental and individual factors that impact education, each factor on the D-STPGE was ranked as either documented and considered as part of the student's unique learning profile or not documented: if such information was not documented, as per best practices in psychological assessment, it was not considered to be important concerning the psychoeducational assessment.

RESULTS

Due to this study being an exploratory frequency distribution on archival psychological reports with the use of the D-STPGE, descriptive statistics were conducted to understand the distribution of the specific construct that was coded and how it relates to the other constructs that were either reported or not reported. Furthermore, a chi-square and the corresponding asymptotic significance value will be analyzed to observe the frequency to which each observed construct was considered and documented.

Descriptive Statistics

In this study, 120 psychoeducational, educational, and neuropsychological reports that focused on disability evaluations for bilingual and multilingual school-aged individuals were coded for developmental factors that appear on the D-STPGE.

Around 69% of the reports coded were initial psychoeducational evaluations that assessed the individual's cognitive, academic, and language functioning to explore any presence of a specific learning disability. 15% of the evaluations coded were mandated triennial evaluations for individuals already receiving special education services while 10% of the evaluations were multi-disciplinary assessments which are neuropsychological (i.e., these evaluations assessed cognitive functioning, academic abilities, language functioning, memory functioning, language dominance, motor skills, adaptive functioning, and so forth). 6% of the evaluations coded were re-evaluations that focused on academic language and speech abilities due to speech delays and the individual's promotion being in doubt. While 81% of these evaluations were on students who were primarily Spanish speakers, around 4% of these evaluations assessed

individuals who were predominantly English speakers. 15% of these evaluations were conducted on individuals who were exposed to English and Spanish and were considered bilingual. Since these evaluations focused on assessing the student's learning, it is important to assess whether such psychologists reported the language achievement tests that are administered to ELs who are learning English in the state of California (these tests happen to be the ELPAC or CELDT). Around 55% of the reports coded reported the student's ELPAC and/or CELDT performance. While 3% of the evaluations were on kindergarten children and 6% on middle school students, most of the evaluations coded were on early elementary school children (around 29% for first graders, 23% for 2nd graders, and 18% for 3rd graders). 8% of the evaluations were on 4th graders while 13% of the evaluations were conducted on 5th graders. Table #1 demonstrates the frequencies of the referral concerns, language of assessment, the student's grade, and reports of language achievement for the data coded.

Table 1. Descriptive Information of Reports Coded

Type of and Reason for Evaluation	**Frequency (N=120)**
Initial Psychoeducational Assessment for Specific Learning Disability	69% (n=83)
Initial Multidisciplinary Assessment for referrals where students presented with many challenges	10% (n=12)
Mandated Triennials Evaluations	15% (n=18)
Re-evaluation due to learning challenges, lack of progress, and/or concerns with Language/Speech	6% (n=7)
Language of the Assessment	**Frequency (N=120)**
Primarily in **Spanish**	81% (n=98)
Primarily in **English**	3% (n=4)
Balanced **(**relatively equal **English** and **Spanish)**	15% (n=18)
Grade	Frequency (N=120)
Kindergarten	3% (n=3)

1st Grade	29% (n=35)
2nd Grade	23% (n=27)
3rd Grade	18% (n=21)
4th Grade	8% (n=10)
5th Grade	13% (n=16)
Middle School	6% (n=8)

Figure 1. Type Of/Reason for Evaluation

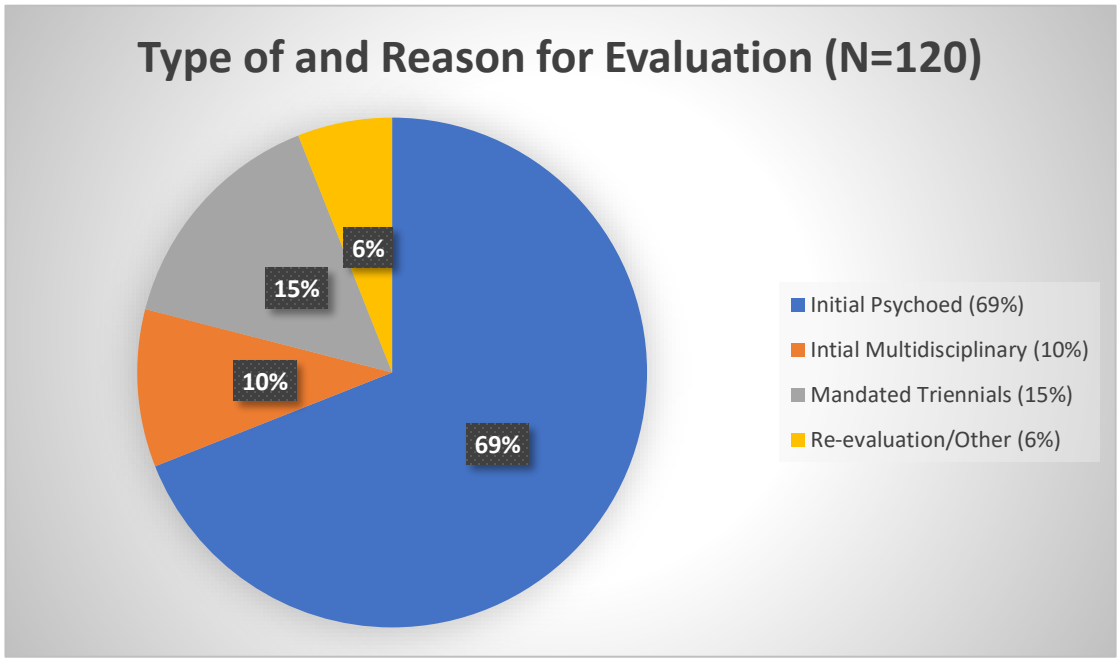


Figure 2. Language of Assessment

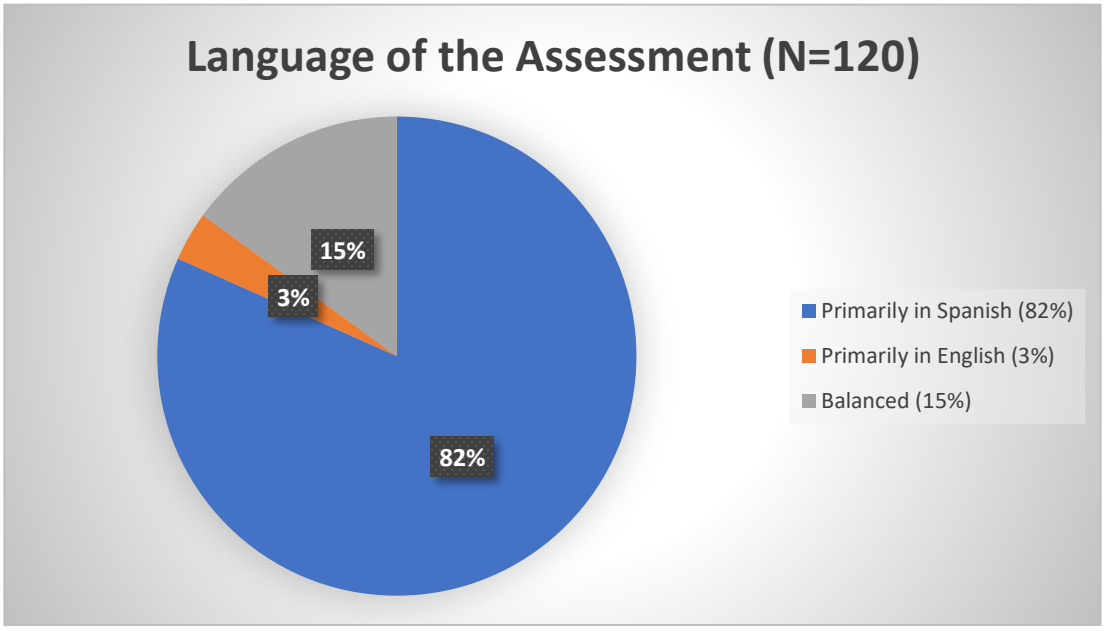


Figure 3. Language Proficiency Test Reported

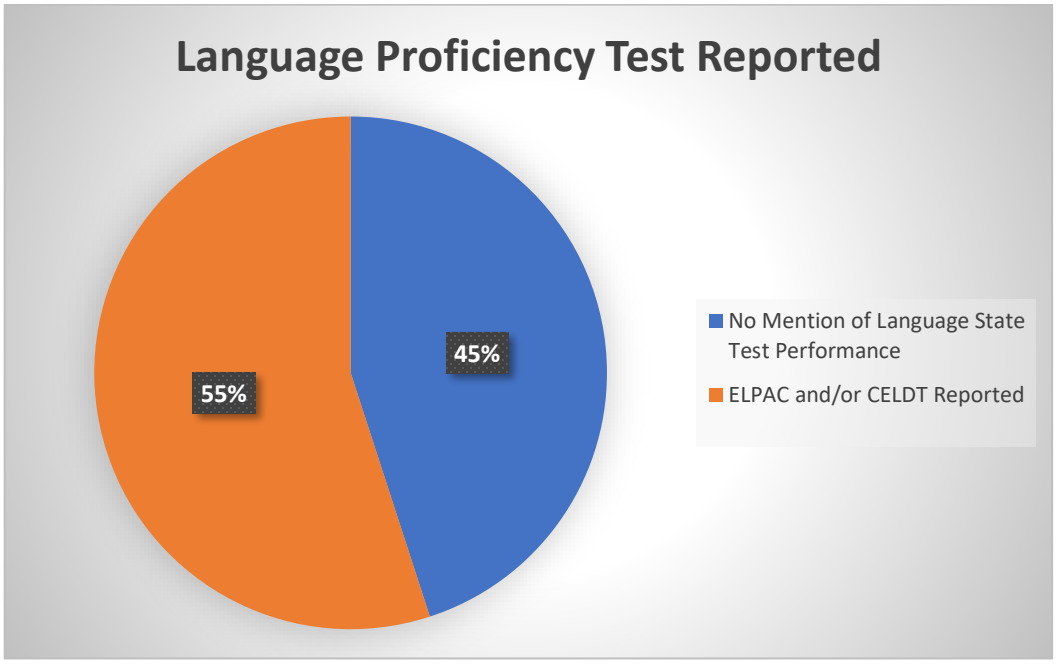
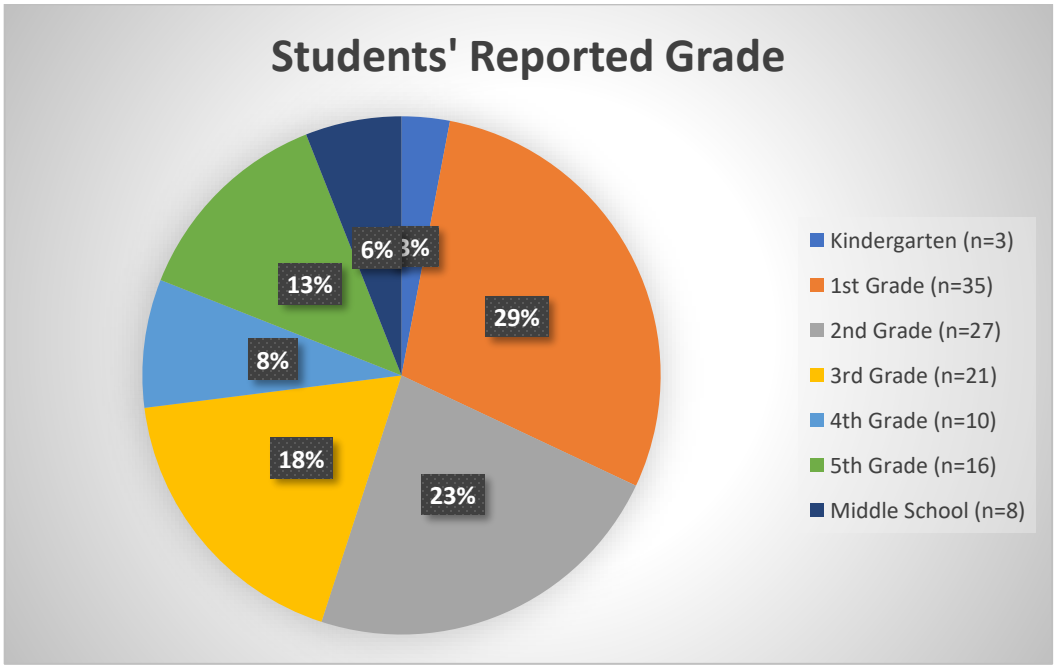


Figure 4. Students Reported Grade



Analyses

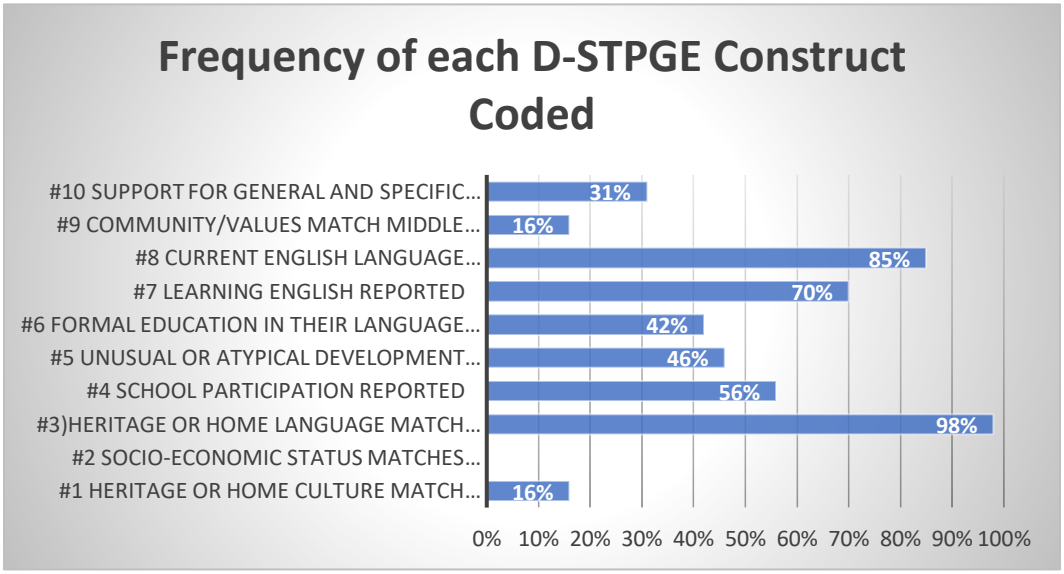
After understanding the nature and language of the assessments coded, a careful analysis of the presence of each developmental factor on the D-STPGE was explored within the data. This information can be seen below in Table 2. When observing the frequency data, it is easy to see constructs that were more likely to be considered in the reports coded, which include Heritage/Home Language (98%), Current English Language Development (85%), and Length of Learning English (70%). Constructs that were less likely to be considered compared to the aforementioned items include School Attendance/Participation (56%), Unusual/Atypical Developmental Life Experience (46%), and Formal Education in Native Language (42%). Constructs that were either not likely to be reported or not reported at all include Support for Social/Emotional Development (31%), Community Language (16%), Heritage/Home Cultural Match Middleclass USA (16%), and SES (0%). The following tables demonstrate the frequency of each item reported in the data.

Table 2. Frequency of each D-STPGE Construct Coded

D-STPGE Items	**Frequency (Sum) of Reported Construct Count: (N=120)**
1) "Heritage or home culture match middle-class, mainstream U.S. values, traditions, and beliefs in supporting the acquisition of school-based acculturative knowledge"	16% (n=20)
2) "Socio-economic status matches middle-class, mainstream families in the U.S. in support in the development of school-based learning and academic skills."	0% (n=1)

3) Heritage or home language match middle-class, mainstream, monolingual English speakers in supporting the acquisition of school-based English?	98% (n=118)
4) "Been able to participate in and attend school consistently and sufficiently to provide an adequate opportunity for academic learning."	56% (n=67)
5) "Avoided any unusual or atypical developmental life experiences (e.g., frequent moves, immigration or migrant worker experience, refugee status, exposure to war, trauma, violence, neglect, abuse, etc.)"	46% (n=55)
6) "Been provided formal education in their heritage language while attending school in the U.S. or before coming to the U.S.?"	42% (n=51)
7) "Been learning English over their lifetime as compared to middle-class, monolingual English speakers of the same age?"	70% (n=84)
8) "Current English language development (including reading/writing, vocabulary, and aspects of speech) compared to middle-class, monolingual, English-speakers of the same age?"	85% (n=102)
9) "Reside in a community where English is the dominant language and where cultural values and attitudes primarily reflect middle-class, mainstream U.S. values?"	16% (n=20)
10) "Been provided support for general and specific aspects of social/emotional development (i.e., mental health, self-identity, gender identification, sexual orientation, self-esteem, etc.)"	31% (n=37)

Figure 5. Frequency as a percentage of evaluations reporting D-STPGE items.



To understand the presence and frequency of each construct on the D-STPGE within the data that was analyzed, a chi-square analysis and its corresponding asymptotic (two-tailed) significance statistic for each item was conducted. A two-tailed test was used given that reporting rates/frequencies could either be much higher than chance expectations (i.e., 50/50) or much lower. With a sample size of 120, the expected level of reporting would therefore be a rate of 60 for both the presence and absence of a construct being reported. The implications regarding the fact that expectations for reporting are not meant to be random or evenly distributed will be presented later in the discussion section. Results and tables that detail the test statistic for each construct coded can be found below.

For the first construct of the D-STPGE, which asks to compare the individual's heritage and home culture to the middle-class USA, a consideration of around 16% was observed (N=20). The chi-square analysis for the Heritage/Home Culture Match Middleclass USA construct (X^2 (1, N = 120) = 53.33, $p < .000$) was found to be statistically significant at the $p < .000$ level. As seen in Table 3, this indicates that this item was less likely to be reported throughout the evaluations that were examined.

Table 3. Chi-Square Analysis of DSTPGE Item #1: Heritage/Home Culture Match Middleclass USA Reported (D-STPGE Item #1)

Reported	Observed N	Expected N	Residual
No	100	60.0	40.0
Yes	20	60.0	-40.0
Total	120		

Test Statistic: Heritage/Home Culture Match	
Chi-square	53.33
Degrees of Freedom	1
Asymptotic Significance	**.000**

The second construct taps into the individual's and their family's socioeconomic status when compared to the middle-class USA. When observing each reported code for this exploratory study, this construct was not likely to be considered across evaluations at all with just a single case found (N=1). A chi-square analysis for this construct demonstrates statistical significance at the .000 level (X^2 (1, N = 120) = 116.03, $p < .000$). As it is evident in Table 4, this result indicates that this item was reported only once throughout all the 120 evaluations that were examined. In other words, the consideration of this construct did not meet the expected statistical frequency for this study.

Table 4. Chi-Square Analysis of DSTPGE Item #2: SES Match Middleclass USA Reported (D-STPGE Item #2)

Reported	Observed N	Expected N	Residual
No	119	60.0	59.0
Yes	1	60.0	-59.0
Total	120		

Test Statistic: Socioeconomic Status Match	
Chi-square	116.03
Degrees of Freedom	1
Asymptotic Significance	**.000**

The third item in the D-STPGE that was evaluated was the extent to which the Heritage/Home Language was reported. Descriptive statistics already confirmed that this

construct was very frequently reported and found in 98% (N=118) of the 120 reports that were evaluated. Subsequent chi-square analysis for this construct demonstrated that the Heritage/Home Language (X^2 (1, N = 120) = 112.133, $p < .000$) was statistically significantly reported at rates that well exceed random reporting. Table 5 provides the results of this analysis which suggests that this item was routinely included in nearly every evaluation that was analyzed in the study.

Table 5. Chi-Square Analysis for D-STPGE Item #3: Heritage/Home Language Reported (D-STPGE Item #3)

Reported	Observed N	Expected N	Residual
No	2	60.0	-58.0
Yes	118	60.0	58.0
Total	120		

Test Statistic: Heritage/Home Language	
Chi-square	112.13
Degrees of Freedom	1
Asymptotic Significance	**.000**

The fourth item of the D-STPGE that was evaluated was School Attendance/Participation and it was found to be reported in 55% (N=67) of the cases. A chi-square analysis for this construct demonstrates that the distribution of reports for the construct of School Attendance/Participation reported (X^2 (1, N = 120) = 1.63, $p = .201$) is essentially random, that is, the result did not suggest statistically significant higher or lower rates of reporting than would otherwise be expected. The results of this analysis are provided in Table 6.

Table 6. Chi-Square Analysis for D-STPGE Item #4: School Attendance/Participation Reported (D-STPGE Item #4)

Reported	Observed N	Expected N	Residual
No	53	60.0	-7.0
Yes	67	60.0	7.0
Total	120		

Test Statistic: School Attendance/Participation	
Chi-square	1.63
Degrees of Freedom	1
Asymptotic Significance	**.201**

The fifth construct of the D-STPGE strives to understand the individual's developmental life experiences and if they experienced anything atypical or unusual. This construct was observed to be reported in 46% of the time (N=55). A chi-square analysis of this construct's reporting demonstrates that Unusual/Atypical Developmental Life Experiences (X^2 (1, N = 120) = 0.833, p = .361) occurred randomly, that is, the resulting frequencies of this reported construct were within the expected statistical frequency range. Table 7 provides the results of this analysis.

Table 7. Chi-Square Analysis for D-STPGE Item #5: Unusual/Atypical Developmental Life Experiences Reported (D-STPGE Item #5)

Reported	Observed N	Expected N	Residual
No	65	60.0	5.0
Yes	55	60.0	-5.0
Total	120		

Test Statistic: Unusual/Atypical Developmental Experiences	
Chi-square	.833
Degrees of Freedom	1
Asymptotic Significance	**.361**

The sixth construct on the D-STPGE that was evaluated concerns the amount or extent of Formal Education in the Native Language. Based on the descriptive statistics, this construct was found to be reported in 42% of the cases (N=51). A chi-square analysis of this item (X^2 (1, N = 120) = 2.70, p = .100) demonstrates that the rate of reporting vs. not reporting is relatively equal and otherwise random and within the frequencies that would be expected for this study. In other words, this construct was not under or over-reported beyond expected statistical frequency levels. Table 8 provides the results of this analysis.

Table 8. Ch-Square Analysis for D-STPGE Item #6: Formal Education in Native Language Reported (D-STPGE Item #6)

Reported	Observed N	Expected N	Residual
No	69	60.0	9.0
Yes	51	60.0	-9.0
Total	120		

Test Statistic: Formal Education in Native Language	
Chi-square	2.70
Degrees of Freedom	1
Asymptotic Significance	**.100**

The seventh construct on the D-STPGE that was observed was how often the Length of Learning English was considered when evaluating a student. This construct was found to be reported somewhat frequently and occurred in 70% of the cases (N=84). A chi-square analysis produced a value that demonstrates that this construct (X^2 (1, N = 120) = 19.20, p < .000) was statistically significant at the p < .000 level. Table 9 provides the results of this analysis which indicates that the Length of Learning English was reported at a rate higher than the expected statistical frequency.

Table 9. Chi-Square Analysis for D-STPGE Item #7: Length of Learning English Reported (D-STPGE Item #7)

Reported	Observed N	Expected N	Residual
No	36	60.0	-24.0
Yes	84	60.0	24.0
Total	120		

Test Statistic: Length of Learning English	
Chi-square	19.20
Degrees of Freedom	1
Asymptotic Significance	.000

The eighth construct from the D-STPGE that was evaluated involved consideration of a student's current English Language Development. Descriptive analysis revealed that this construct was reported at a very high frequency of 85% (N=102). A chi-square analysis produced values (X^2 (1, N = 120) = 58.80, $p < .000$) that were found to be statistically significant at the $p < .000$ level. This suggests that English Language Development was reported very frequently and in nearly all the cases examined. Table 10 provides the results of this analysis which suggests that English Language Development was reported at very high rates beyond the expected statistical frequency for this study.

Table 10. Chi-Square Analysis for D-STPGE Item #8: Current English Language Development Reported (D-STPGE Item #8)

Reported	Observed N	Expected N	Residual
No	18	60.0	-42.0
Yes	102	60.0	42.0
Total	120		

Test Statistic: Current English Language Development	
Chi-square	58.80
Degrees of Freedom	1

Asymptotic Significance	.000

The ninth construct from the D-STPGE that was examined was consideration of a student's Community Language which is distinct from the heritage or home language. According to the descriptive analysis, this item was reported in only 17% of the cases (N=20). A chi-square analysis on this construct demonstrated that the result was statistically significant at the $p < .000$ level (X^2 (1, N = 120) = 53.33, $p < .000$). Results of this analysis are provided in Table 11 and indicates that Community Language was reported at a lesser frequency than expected and did not reach the expected statistical frequency for this study.

Table 11. Chi-Square Analysis for D-STPGE Construct Item #9: Community Language Reported (D-STPGE Item #9)

Reported	Observed N	Expected N	Residual
No	100	60.0	40.0
Yes	20	60.0	-40.0
Total	120		

Test Statistic: Community Language	
Chi-square	53.33
Degrees of Freedom	1
Asymptotic Significance	.000

The tenth and final construct from the D-STPGE that was examined for this study involved consideration of the extent of support for various social/emotional factors that can affect development. According to the descriptive statistics, this item was reported in 31% of the cases evaluated (N=37). A chi-square analysis for this item (X^2 (1, N = 120) = 17.63, $p < .000$) demonstrated that the reporting rate was statistically significant at the $p < .000$ level. This suggests that the degree to which social or emotional factors and their

potential impact on development was statistically underreported at frequencies much lower than would otherwise be expected. The results of this analysis are provided in Table 12.

Table 12. Chi-Square Analysis for D-STPGE Item #10: Support for Social/Emotional Development Reported (D-STPGE Item #10)

Reported	Observed N	Expected N	Residual
No	83	60.0	23.0
Yes	37	60.0	-23.0
Total	120		

Test Statistic: Social-Emotional Development	
Chi-square	17.63
Degrees of Freedom	1
Asymptotic Significance	.000

DISCUSSION

The purpose of this study was to gain a better understanding of the consideration of a wide range of developmental factors related to English Learners that would set them distinctly apart from the development of monolingual, mainstream English speakers. Such differences, and more importantly the magnitude of the difference, are crucial factors in providing the appropriate context by which to interpret standardized test scores, particularly for disability identification. The question, of course, was whether and to what extent these factors are being considered in disability evaluations of English learners currently being conducted.

To address this question, a sample of 120 de-identified reports of psychoeducational evaluation were carefully reviewed to see whether the various elements of developmental differences were examined, considered, and reported. The developmental factors were based on the D-STPGE (Ortiz, 2023) which is a semi-structured type of interview that examines 10 different developmental domains and provides a mechanism for rating the differences relative to monolingual, mainstream, English speakers. Its purpose is to assist practitioners in being able to establish how similar (or different) a culturally and linguistically diverse student might be as compared to the general, English-speaking, mainstream population on which tests are typically normed.

The expectation within this study was that all the elements in the D-STPGE would be observed and likely reported at very high rates given their importance in being able to interpret data fairly and in conducting nondiscriminatory assessment. Results from analyses of the rates and frequency of reporting (i.e., documenting) consideration of all

10 elements contained in the D-STPGE provide a window into the extent to which evaluators (bilingual and otherwise) are or are not documenting consideration of these factors and their potential impact on development which would, in turn, affect the interpretation of any measure of intelligence, academic achievement, and language functioning. The results indicated that some constructs, particularly those related to English language development, were likely to be considered routinely whereas other factors tended to be exceedingly ignored and underreported, especially within the context of what is supposed to be a nondiscriminatory evaluation.

The chi-square results demonstrated that the following constructs were considered at statistically much higher rates and frequencies than what would be expected merely by chance: Heritage/Home Language (Item #3: 98%), Current English Language Development (Item #8: 85%), and Length of Learning English (Item #7: 70%). Overall, evaluators recognize the importance of language development and that they are quite reliable in highlighting it within the context of their findings and reports. This makes perfect sense and is a sign of good practice, however, note that not all evaluators are considering language because the rates are not 100% as might be expected for any evaluation that would meet best practice standards. Language development differences are rather obvious factors to document within a bilingual evaluation and many evaluators have learned throughout their training that such issues are imperative to explore and note. The fact that these constructs tended to be documented at a much higher rate as compared to the other constructs within the D-STPGE indicates what might be a too narrow focus on language regarding bilingual evaluations and concomitant lack of sufficient consideration of other developmental and environmental factors.

Conversely, four constructs were statistically far less likely to be reported despite also being important components to consider when evaluating diverse learners. Constructs that were less likely to be reported and below the expected statistical frequency included: Support for Social-Emotional Development (#10: 31%); Home Culture Match to U.S. Middle Class (Item #1: 16%); Community Language (Item #9: 16%); and Socio-economic Status match to U.S. Middle Class (Item #2: 0%). These constructs were observed to be statistically far below what would otherwise be expected due to chance (50%), but they are also far below clinical expectations for evaluations involving culturally and linguistically diverse students wherein every such evaluation should include them. Variables related to socio-economic status, heritage cultural development, exposure to a minority language in the community, and support for social-emotional challenges are all critical factors that must be considered in an evaluation given how much they can affect cognitive, linguistic, and academic development and measured performance. Not only was the frequency of reporting very low for these factors, but they were also exceptionally low relative to clinical expectations. Put another way, they were effectively ignored in 69%, 84%, 84%, and 99% of all cases. Socio-economic status and its match to the middle class was an area in which practically every evaluation failed to assess its importance which is surprising given that it is a specific and noted exclusionary factor for any evaluation conducted to identify Specific Learning Disability. That it was conspicuously absent from the evaluations is cause for alarm and accompanied by the poor rates noted for the reporting of the other variables, strongly suggests that evaluators are failing dramatically in considering these four critical elements of a nondiscriminatory evaluation. If the sample of 120 evaluations used in this

study is representative of typical evaluation practices, then taken together, they demonstrate an overwhelmingly poor level of compliance with reporting and documenting procedures necessary to ensure fair and nondiscriminatory evaluation of bilingual and culturally diverse pupils.

The results also noted that there were three constructs found to be reported at rates that would be expected randomly, or at least by chance (50/50) from a statistical perspective. These included: School Attendance/Participation (Item #4: 56%), Unusual or Atypical Development (Item #5: 46%), and Formal Education in their Native Language (Item #6: 42%). According to the results, these elements were reported about as often as they were unreported meaning that there was no evidence of over or under-reporting noted in the evaluations. However, from a clinical perspective, the expectation is that they should have been routinely and frequently reported, much like all the other developmental variables outlined in the D-STPGE. Lack of school attendance or participation, atypical developmental experiences, and the extent of formal education in a child's heritage language are not factors that should be hit or miss in terms of their inclusion in an evaluation. Failure to properly consider any one or all these factors re important when considering whether a diverse student's performance on a variety of tests is valid or invalid. Much as with socio-economic status, lack of school attendance represents another exclusionary factor in the identification of Specific Learning Disability and merits specific consideration. The presence of unusual or atypical developmental experiences as well as the provision of (or lack thereof) formal education in a student's heritage language can both have powerful effects on the development of diverse children which is readily measurable by standardized tests, particularly those focusing on

cognitive ability acquisition of academic skills. Thus, although these developmental factors were neither over nor under-reported statistically speaking, the fact remains that they should have been over-reported—that is, they should have been routine and typical considerations documented in any evaluation of culturally and linguistically diverse individuals, particularly those suspected of having a Specific Learning Disability (SLD).

This study also presents the importance of bilingual and nondiscriminatory testing that goes beyond psychometrics and standardized testing. Many psychologists in the field have been shedding light on different testing measures for multilingual and multicultural individuals since comparing their performances to standard measures has been considered discriminatory for many decades. Instruments like X-BASS (which includes the C-LIM, C-LIM+ATE, and D-STPGE) can help clinicians understand a bilingual student's performance on cognitive measures and compare them to how typical ELs perform on such measures. This includes the exact qualitative variables that were highlighted and analyzed throughout this study. Furthermore, a scale like the D-STPGE can guide clinicians in understanding possible developmental factors when assessing ELs that deviate from other ELs or mainstream pupils. Such a scale can help quantify this difference and compare it to how individuals with similar development perform on cognitive measures. Such practices will help bolster the validity of a psychoeducational evaluation.

The fact that California has the highest EL population in the U.S. and that much of the school-age population comes from diverse cultural and linguistic backgrounds, coupled with the State's history of embracing multilingualism and multiculturalism, makes a case for how shocking the results are given that these are evaluation completed

in a state where these factors exist. This might not be as surprising in states where diversity is less, but it means that even in a state where such developmental factors should be considered, they aren't being considered well, or as much as would be expected.

Limitations

Although the study included a relatively large number of evaluations, there are still some limitations that may have affected the generalizability of the findings, including the limited geographical location of the sample. Although 120 reports were included in the study, all of them were from the state of California. The results of this study would likely be more generalizable if reports from other states including those with a high relative population of ELs (i.e., New York, Massachusetts, Florida, Texas, and so forth). If such potential difficulties exist in the current practice where there exists a high percentage of bilingual and multicultural individuals, many problems can arise with evaluations in these states as well as states with lower percentages of ELs and diverse students. A larger sample considering the aforementioned factors may have helped further analyze which developmental factors tend to be reported more. Other limitations of this study include the psychometric properties of the D-STPGE. This study is a pilot study to understand which constructs, if any, and to what extent, school psychologists are considering in nondiscriminatory assessments. Further studies and research are needed to understand the validity that the D-STPGE holds.

Future Directions

This study focused on coding 120 psychological reports that strived to evaluate bilingual and multilingual developing children from kindergarten to 6th grade. Throughout the analysis of the data with the use of the D-STPGE to code for developmental factors, it was found that many evaluators were not considering a significant number of such factors that are important to consider when conducting a nondiscriminatory psychological assessment. Highlights of this research demonstrate that many evaluators tend to focus on factors that are very important for evaluations but often straightforward when one is conducting a bilingual evaluation (i.e., home language, current language development, and length of learning English). However, many factors that tended to not be considered were factors that were less obvious to an evaluator. Such factors include atypical developmental experiences, heritage/home culture, community values, and so forth. Constructs like school participation were less likely to be reported, demonstrating weaknesses in the evaluator's consideration for significant qualifying factors for specific instructional services like special education. Further research and integration on developmental differences that exist amongst ELs are important to consider for the training of school psychologists. Future directions for this study include further research on the validity and reliability of the D-STPGE as it is used to understand multilingual and multidisciplinary evaluations within the scope of nondiscriminatory assessments. Other directions of this study include understanding the correct administration of the D-STPGE when evaluating a bilingual pupil.

Implications for School Psychologists

Currently, the field of psychometrics lacks different methods available to an evaluator that helps them quantify the differences that exist amongst ELs. The D-STPGE is a scale that is a step forward to fixing this discriminatory issue that exists in the field. Although certain constructs were likely to be reported, many others were less likely, if at all, considered when evaluating bilingual students. A scale like the D-STPGE can guide clinicians by providing them with an estimate and comparison of an EL's development difference to other ELs. Overall, school psychologists who conduct bilingual and multilingual/nondiscriminatory assessments need a guide like the D-STPGE to be able to compare and quantify the difference that exists between that EL amongst others within his respective peer group.

Due to the results of this study, more teaching and training are needed to bring these diverse issues to the forefront of evaluations. This includes factors that were found in this study to be less likely to be reported like SES, school participation, unusual/atypical development, formal education in heritage language, support for social-emotional development, as well as the community values concerning a young child's cognitive and academic development. Furthermore, the results of this study support a change in report writing requirements when one evaluates an EL or any other individual who is multicultural. A psychological report is not only a legally binding document, but it essentially follows the child throughout development; if those downstream who read the report are not aware on what basis such decisions were made, it undermines the consistency in terms of support and the credibility of the original

evaluation. Furthermore, these issues correlate in part with the Individuals with Disabilities Education Act (IDEA), such as the exclusionary factors for SLD. This implies that there are legal requirements for the inclusion and consideration of developmental factors in any psychological or psychoeducational evaluation. Yet, whether these factors are being considered is not evident in the findings of this study, let alone how consideration was given to determining whether they were non-factors, contributory factors, or primary factors affecting the student's test performance and schoolwork. Thus, these findings could be used to promote specific standards for use when reporting the results of an evaluation with Els and should be integrated within the psychoeducational report writing process. Taken together, it is not possible to defend the results of an evaluation if such developmental factors are not considered when evaluating bilingual and multilingual learners.

It is imperative to paint an accurate picture of a bilingual child; this entails comparing the child to others like the child and not necessarily the mainstream culture if this child's cultural and linguistic development is distinct. Future steps should include updating standards for bilingual evaluation which involves assessing developmental factors that aid in or inhibit cognitive and academic development. Training on cross-battery assessments and evaluations that include bilingual norms should also be considered and highlighted during a bilingual school psychologist's training and practice.

REFERENCES

Alfonso, V. C., Bracken, B. A., Nagle, R. J., Ortiz, S. O., & Wong, J. Y. T. (2020). Psychoeducational Assessment of Culturally and Linguistically Diverse Children. In *Psychoeducational Assessment of Preschool Children*. essay, Routledge.

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.

August, D., & Haynes, E. (2016, April). What Will ESSA Mean for English Learners? https://www.air.org/resource/what-will-essa-mean-english-learners.

Burr, E., Haas, E., & Ferriere, K. (2015). *Identifying and supporting English learner students with learning disabilities: Key issues in the literature and state practice* (REL 2015–086). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from: http://ies.ed.gov/ncee/edlabs.

Berman, A. I., Haertel, E. H., & Pellegrino, J. W. (Eds.). (2020). Comparability of Large Scale Educational Assessments: Issues and Recommendations. Washington, DC: National Academy of Education.

Blitz, L. V., Anderson, E. M., & Saastamoinen, M. (2016). Assessing perceptions of culture and trauma in an elementary school: Informing a model for culturally responsive trauma-informed schools. *The Urban Review*, *48*(4), 520-542

Chu, S.-Y., & Flores, S. (2011). Assessment of English language learners with learning disabilities. *Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, *84*(6), 244–248. http://eric.ed.gov/?id=EJ950410

Clark-Gareca, B. (2016). *Pressure to Pass: NYSITELL and NYSESLAT Tests*, *3*(1), 3540. journal.nystesol.org.

Clark-Gareca, B., Short, D., Lukes, M., & Sharp-Ross, M. (2019). Long-term English learners: Current research, policy, and practice. *TESOL Journal*, *11*(1). https://doi.org/10.1002/tesj.452

Cormier, D.C., McGrew, K.S. & Ysseldyke, J. E. (2014). The Influences of Linguistic Demand and Cultural Loading on Cognitive Test Scores. Journal of Psychoeducational Assessment, 32(7), 610-623.

Cormier, D. C., Bulut, O., McGrew, K. S. & Kennedy, K. (2022). Linguistic Influences on Cognitive Test Performance: Examinee Characteristics Are More Important than Test Characteristics, Journal of Intelligence, Volume 10, Issue 1.

Delgado, C. E. F., & Scott, K. G. (2006). Comparison of referral rates for preschool children at risk for disabilities using information obtained from birth certificate records. Journal of Special Education, 40, 28 35.doi:10.1177/00224669060400010301

DeMatthews, D. E., Edwards, D. B., & Nelson, T. E. (2014). Identification problems: US special education eligibility for English language learners. *International Journal of Educational Research, 68*, 27–34. https://doi.org/10.1016/j.ijer.2014.08.002

Department of Education, U. S. (2017). *Revised State Template for the Consolidated State Plan the Elementary and Secondary Education Act of 1965, as amended by Every Student Succeeds Act*. https://www.ed.gov/essa?src=rn. https://www.regents.nysed.gov/common/regents/files/A1%202a%20Revised%20State%20Template%20for%20the%20Consolidated%20State%20Plan_1.pdf.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C. (2015). Cross-Battery Assessment Software System (X-BASS). Hoboken, NJ: John Wiley.Fisher, D. & Frey, N. (2014) Student and Teacher Perspectives on a Close Reading Protocol, Literacy Research, and Instruction, 53:1, 25-49

Figueroa, R., & Hernandez, S. (2000). *Testing Hispanic students in the United States: Technical and policy issues*. Washington, DC: Presidents' Advisory Commission on Educational Excellence for Hispanic Americans.

Freeman, D., & Freeman, Y. (2004). Essential linguistics: What you need to know to teach reading, ESL, spelling, phonics, and grammar. Portsmouth, NH: Heinemann

García, O., Kleifgan, J. A., & Falchi, L. (2008, January). *From English Language Learners to Emergent Bilinguals*, *1*(1), 1–61. Retrieved from https://files.eric.ed.gov/fulltext/ED524002.pdf

Harry, B., & Klingner, J. K. (2006). Why are so many minority students in special education? Understanding race & disability in schools. New York, NY: Teachers College Press

Hesson, S. (2013, September). *Theory, Research, and Action in Urban Education*. Theory, Research, and Action in Urban Education - An Online, Open-Access, Peer-Reviewed Journal Hosted by Urban Education @ The CUNY Graduate Center. https://traue.commons.gc.cuny.edu/issue-2-fall-2013/hesson/

Hesson, S. (2013). *Rethinking Assessment Policies for Emergent Bilinguals in New York State*, *3*(1). Retrieved November 2021, from https://traue.commons.gc.cuny.edu/issue-2-fall-2013/hesson/.

Institute of Medicine and National Research Council. 2015. *Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation*. Washington, DC: The National Academies Press. https://doi.org/10.17226/19401.

Jensen, A. R. (1976). Construct validity and test bias. Phi Delta Kappan, 58, 340–346.

Johnson, A. (2020). The Impact of English Learner Reclassification on High School Reading and Academic Progress. Educational Evaluation and Policy Analysis, 42(1), 46–65. https://doi.org/10.3102/0162373719877197

Kreskow, Kelly (2013). Overrepresentation of Minorities in Special Education. Education Masters. Paper 257. 1-49. https://fisherpub.sjf.edu/education_ETD_masters

Lohman, D. F., Korb, K. A., & Lakin, J. M. (2008). Identifying Academically Gifted English Language Learners Using Nonverbal Tests: A Comparison of the Raven,

NNAT, and CogAT. *Gifted Child Quarterly*, *52*(4), 275-296. https://doi.org/10.1177/0016986208321808

López, Ruby, and Linn, Diana (2018) "Representation of English Language Learners in Special Education: A Campus-level Study," Journal of Multicultural Affairs: Vol. 3: Iss. 2, Article 4. MacSwan, J., & Rolstad, K. (2006). How language proficiency tests mislead us about ability: Implications for English language learner placement in special education. *Teachers College Record*, *108*(11), 2304–2328. http://eric.ed.gov/?id=EJ746179

Morgan, P. L., Farkas, G., Hillemeier, M. M., Mattison, R., Maczuga, S., Li, H., & Cook, M. (2015). Minorities Are Disproportionately Underrepresented in Special Education: Longitudinal Evidence Across Five Disability Conditions. Educational Researcher, 44(5), 278–292. https://doi.org/10.3102/0013189X15591157

New York State Education Department. (2019). *New York State English as a Second Language Achievement Test (NYSESLAT)*. Retrieved from http://www.nysed.gov/state-assessment/new-york-state-English-second language-achievement-test-nyseslat

New York State United Teachers (NYSUT) (2015). Webinar on NYSESLAT, June 9, 2015.

Office of Superintendent of Public Instruction (OSPI) (2015). Washington English Language Proficiency Assessment. Retrieved from http://www.k12.wa.us/Assessment/EL/

Norfolk, P. A., Farmer, R. L., Floyd, R. G., Woods, I. L., Hawkins, H. K., & Irby, S. M. (2015). Norm Block Sample Sizes: A Review of 17 Individually Administered Intelligence Tests. *Journal of Psychoeducational Assessment*, *33*(6), 544–554. https://doi.org/10.1177/0734282914562385

Ortiz, S.O. (2017). Technical manual: Ortiz Picture Vocabulary Acquisition Test (PVAT). North Tonawanda, NY: Multi-Health Systems.

Ortiz, S. O., Ortiz, J. A., & Devine, R. I. (2016). Assessment of Culturally and Linguistically Diverse Individuals with the Woodcock-Johnson IV. In *WJ IV clinical use and interpretation: scientist-practitioner perspectives* (pp. 317–353). essay, Academic Press.

Ortiz, S. O., Piazza, N., Ochoa, S. H., & Dynda, A. M. (2018). Testing with culturally and linguistically diverse populations: New directions in fairness and validity. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 684–712). The Guilford Press.

Ortiz, S. O., & Cehelyk, S. K. (2023, December 31). *The bilingual is not two monolinguals of same age: Normative Testing Implications for multilinguals*. MDPI. https://doi.org/10.3390/jintelligence12010003

Reddy, A. (2008). *The Eugenic Origins of IQ Testing: Implications for Post-Atkins Litigation. Volume 57, Issue 3, Spring 2008 Symposium: Protecting a National Moral Consensus: Challenges in the Application of Atkins v. Virginia. https://via.library.depaul.edu/law-review/vol57/iss3/5*

Rhodes, R., Ochoa, S. H. & Ortiz, S. O. (2005). Comprehensive Assessment of Culturally and Linguistically Diverse Students: A Practical Approach. New York: Guilford.

Robinson-Cimpian, J. P., & Thompson, K. D. (2015). The effects of changing test-based policies for reclassifying English learners. Journal of Policy Analysis and Management, 35(2), 279–305. http://doi. org/10.1002/pam.21882.

Rueda, R., & Windmuller, M. (2006). English language learners, LD, and overrepresentation: A Multiple Level Analysis. Journal of Learning Disabilities, 39(2), 99-107

Santibañez, L., & Gándara, P. (2018, February). *Teachers of English language Learners in Secondary Schools: Gaps in Preparation and Support*. https://files.eric.ed.gov/fulltext/ED586360.pdf.

Salvia, J., & Ysseldyke, J. E. (1991). *Assessment*. Houghton Mifflin Co.

Skiba, R. J., Poloni-Staudinger, L., Simmons, A. B., Feggins-Azziz, L. R., & Chung, C. (2005). Unproven links: Can poverty explain ethnic disproportionality in special education? Journal of Special Education, 39, 130–144. doi:10.1177/00224669050390030101

Shore, J. R., & Sabatini, J. (2009). English Language Learners with Reading Disabilities: A Review of the Literature and the Foundation for a Research Agenda. *ETS Research Report Series*, *2009*(1), i-48. https://doi.org/10.1002/j.2333-8504.2009.tb02177.x

Sotelo, D. M., Ortiz, S. O., Flanagan, D. P., & Chaplin, W. F. (2013). English Language Proficiency and Test Performance: An Evaluation of Bilingual Students with the Woodcock-Johnson III Tests of Cognitive Abilities. Psychology in the Schools, 50(8), 781–797. https://doi-org.jerome.stjohns.edu/10.1002/pits.21706

Sullivan, A. L. (2011). Disproportionality in Special Education Identification and Placement of English Language Learners. Exceptional Children, 77(3), 317–334. https://doi.org/10.1177/001440291107700304

U.S. Department of Education, National Center for Education Statistics, Common Core of Data (2020). Local Education Agency Universe Survey, 2000–01 through 2016–17. Retrieved from: https://nces.ed.gov/programs/digest/d18/tables/dt18_204.20.asp?current=yes

Li, Chieh & Vasquez-Nuttall, E. & Dynda, A. & Ortiz, Samuel & Armengol, Carmen & Walton, J. & Phoenix, K. (2007). Cognitive Assessment of Culturally and Linguistically Diverse Students.

Wolfram, W., Adger, C. T., & Christian, D. (1999). Dialects in Schools and Communities. Mahwah, NJ: Lawrence Werlbaum Associates.

Zhaohui Sheng, Yanyan Sheng & Christine J. Anderson (2011) Dropping out of School among ELL Students: Implications to Schools and Teacher Education, The Clearing House: A Journal of Educational Strategies, Issues, and Ideas, 84:3, 98103, DOI: 10.1080/00098655.2010.538755

Zehler, A. M., Fleischman, H. L., Hopstock, P. J., Pendzick, M. L., & Stephenson, T. G. (2003). *A descriptive study of services to LEP students and LEP students with*

disabilities (No. 4 Special topic report: findings on special education LEP students). Development Associates, Inc.: Arlington, VA.

Vita

Name	*Erick Bienvenido Díaz*
Baccalaureate Degree	*Bachelors of Arts, College Of the Holy Cross, Worcester, MA Major: Psychology & Spanish with a concentration in Latino Studies*
Date Graduated	*May 2014*
Other Degrees and Certificates	*Master of Science, St. Johns University, Queens, NYS, Major: School Psychology*
Date Graduated	*May 2021*