Assessing the Clinical Accuracy of the Illinois Early Intervention Eligibility Criteria

Sarah A. Thomas
Allison Gladfelter

Follow this and additional works at: https://huskiecommons.lib.niu.edu/studentengagement-honorscapstones

Recommended Citation
https://huskiecommons.lib.niu.edu/studentengagement-honorscapstones/192

This Article is brought to you for free and open access by the Undergraduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Honors Capstones by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
NORTHERN ILLINOIS UNIVERSITY

Assessing the Clinical Accuracy of the Illinois Early Intervention Eligibility Criteria

A Capstone Submitted to the

University Honors Program

In Partial Fulfillment of the

Requirements of the Baccalaureate Degree

With Honors

Department Of

Health and Human Sciences

By

Sarah Thomas

Dr. Allison Gladfelter

DeKalb, Illinois

May 2021
Capstone Title (print or type): Assessing the Accuracy of the Illinois Early Intervention Eligibility Criteria

Student Name (print or type): Sarah Thomas

Faculty Supervisor (print or type): Allison Gladfelter

Faculty Approval Signature: Allison Gladfelter

Department of (print or type): Health and Human Sciences

Date of Approval (print or type): December 5th, 2020

Date and Venue of Presentation: Honors Capstone Symposium Virtual Presentation

Check if any of the following apply, and please tell us where and how it was published:

☐ Capstone has been published (Journal/Outlet):

☐ Capstone has been submitted for publication (Journal/Outlet):

Completed Honors Capstone projects may be used for student reference purposes, both electronically and in the Honors Capstone Library (CLB 110).

If you would like to opt out and not have this student’s completed capstone used for reference purposes, please initial here: _______ (Faculty Supervisor)
Abstract

**Purpose:** To qualify for Early Intervention in Illinois, a child must show a 30% developmental delay based on age-equivalent (AE) scores. Unfortunately, the use of AE scores is not empirically supported for determining intervention eligibility because they do not account for the normal variability of scores found within a typical population. The purpose of this project was to establish the likelihood of eligibility misclassifications using the Illinois criterion based on AE scores.

**Method:** Scoring tables from three Illinois pre-approved tests were digitally converted and analyzed using a code written in Visual Basic for Applications, the macro-programming language in Excel, and Python. This code compared all possible combinations of standard scores, AE scores, raw scores, and chronological ages from birth to 36 months to calculate the likelihood of misclassifications based on the Illinois eligibility criterion.

**Results:** In all three tests, misclassifications were identified. In total, 2,447 combinations of scores were found that could lead to the denial of services to children who exhibit developmental delays and 2,966 combinations that could lead to the inappropriate provision of services to typically developing children.

**Conclusions:** The Illinois eligibility criterion based on AE scores was found to be problematic, as shown by the thousands of combinations of scores that could lead to misclassifications on these three tests alone. Early Intervention eligibility criteria must be modified to reflect the best available evidence and in accordance with the administration and scoring procedures of the testing manuals to ensure that these critical services reach the children who need them.
Introduction

Under the Individuals with Disabilities Education Act (1997), each state is mandated to enact its own eligibility criteria to determine who qualifies for Early Intervention services. Illinois requires a child to show a 30% delay in development through the use of age equivalent (AE) scores from pre-approved diagnostic tools (IDHS, 2016). Using this eligibility criterion, children who are not developing at an expected rate are provided with Early Intervention services. Although the provision of Early Intervention has a myriad of benefits for children at risk for developmental delays, the use of AE scores, rather than standard scores, as a diagnostic indicator has not been scientifically supported. A standard score is recommended because it incorporates the normal variability of scores within a normal population distribution and compares an individual’s distance from the mean and standard deviation from that normal distribution (Andersson, 2004). In other words, it directly compares a child’s performance to the performance of chronological age-matched peers. Alternatively, an AE score does not directly compare a child’s score to their chronological age-matched peers but instead reflects the age cohort that the child's raw score is equal to the mean or median score. As such, an AE score does not provide any information regarding whether the child’s performance is within the typical range of scores expected for their age. Because of this limitation, the Illinois Early Intervention eligibility criterion that relies on AE scores puts children at a greater risk of misclassification, which could result in the inappropriate provision of services.

Misclassifications can lead to a plethora of problems. First and foremost, they can prevent children from receiving the services they need. If children do not receive the proper services at a young age, it can be detrimental to their future (Karoly, 2005). Early Intervention has long-lasting impacts, such as increases in cognitive and behavioral development that
continue into adulthood (Karoly, 2005). Additionally, Early Intervention affects children’s social, emotional, physical, and mental health, all of which have long-term benefits (Karoly, 2005). Because of these well-established benefits, families of children who need services frequently advocate for Early Intervention services for their children (CPIR, 2012). For these reasons, intervention eligibility decisions must be based on the best available evidence to correctly identify those most in need.

Alternatively, the inappropriate provision of services based on un-scientific eligibility criteria harms those beyond the families of children with developmental delays. Taxpayers pay for Early Intervention services, which is a government-funded program. The majority of Early Intervention program funding comes from General Revenue funds and from the department of Health Care and Family Services (HFS) through Medicaid reimbursement (IDHS). In 2016, the Early Intervention expenditures in Illinois totaled $161,800,000 (Kids Count Data Center). Accurate distribution of Early Intervention services is essential for all Illinois taxpayers, not only the families of children with developmental delays. Taxpayers must have confidence that their money is going toward the appropriate provision of services.

As such, the purpose of this project was to assess the likelihood of misclassifications using the Illinois Early Intervention eligibility criterion based on AE scores. To establish these classification likelihoods, scoring data were extracted from administration manuals selected from the Illinois Early Intervention Approved Evaluation and Assessment Instruments list to analyze all possible combinations of age, raw scores, AE scores, and standard scores. Using the Illinois Early Intervention 30% developmental delay criterion, these numbers were used to calculate classification likelihood ratios and then compared to the standard scoring guidelines recommended within each test. Prior research has extensively reported the limitations of using
AE scores to identify developmental delays (as summarized in Andersson, 2004). This study extends this previous research by unveiling the likelihood of misclassifications that could lead to the inappropriate provision of Early Intervention services because of the mandated use of AE scores in the state of Illinois.

**History of Early Intervention Laws**

To gain more insight into Early Intervention, it is important to look at its origin. In 1975 Congress passed Public Law 94-142, *Education for All Handicapped Children Act*, which was later changed to the *Individuals with Disabilities Education Act* (IDEA) in 1990. Under this four-part legislation, children with disabilities must legally be provided with services tailored to their needs so that they can access the same opportunities as a child who does not have a disability. Between 1996-2004, Congress passed Public Law 108-446, or Part C of IDEA, which mandated that state and local agencies must provide families with services regardless of their socio-economic status and race (CPIR, 2012). The federal IDEA Part C law allows each individual state to determine its own criteria for service eligibility.

**Qualifying for Early Intervention Services in Illinois**

In Illinois, there are two eligibility criteria set in place for a child to receive Early Intervention services. The first qualifying criterion is if the child is medically diagnosed with a known physical or mental condition (Berman, 2016). A child diagnosed with Down Syndrome, Cerebral Palsy, Rett Syndrome, or any other medical condition would qualify under this criterion. The second eligibility criterion is if a child experiences a 30% developmental delay, based on an AE score, in one or more of the following areas: cognitive, physical, social-emotional, communication, or adaptive development (Berman, 2016). Percent developmental delay is calculated by taking the chronological age (in months) and subtracting the age
equivalence (in months), then dividing by the chronological age (in months), and finally multiplying by 100 (see Equation 1 below).

\[
\text{Equation 1. } \frac{\text{Chronological age - Age equivalence}}{\text{Chronological age}} \times 100 = \text{Percent Developmental Delay}
\]

If the percent developmental delay is greater than 30%, then the child is considered to have a developmental delay according to the Illinois eligibility criterion. For example, using scoring data from the Preschool Language Scale-5th Edition (PLS-5; Zimmerman et al, 2011), a pre-approved assessment instrument in Illinois, a 35-month-old child with a raw score of 26 on the Auditory Comprehension scale would receive an age equivalent score of 23 months. Her percent developmental delay for Auditory Comprehension would then be calculated as follows:

\[
\frac{35 - 23}{35} \times 100 = 34\%
\]

Based on the Illinois 30% developmental delay eligibility criterion, this child would be classified as performing below age expectations and would be eligible to receive Early Intervention services.

**Obtaining Early Intervention Services in Illinois**

Multiple steps must be taken for a child to receive services when Illinois families express concern for their child. First, a family could receive a formal referral to Child and Family Connections, an Illinois-based program that provides Early Intervention services for children with developmental delays. Once a referral is made, a meeting with case managers or social workers is held with the families so that they can better understand what Early Intervention services may be appropriate. IDEA requires that providers make referrals within “7 days after the infant or toddler is identified as having a possible delay or disability.” (ASHA, n.d.). The provider has 45 days to complete screening which includes interviews, exploring evaluations, screening for eligibility, and completion of initial paperwork. From there, evaluations and assessments are given by professionals such as speech-language pathologists, occupational
therapists, physical therapists, social workers, psychologists, or developmental therapists. Next, the family and case managers work to establish the Individualized Family Services Plan (IFSP), which details the services for the child and is mandated by IDEA. Once the child is deemed eligible for services, the family is connected to service providers. Within six months, the service providers give annual reviews and transition into the program. At any point during the process, the family can be denied or may choose to withdraw from services. Then, they would start the process over if they later wanted to again receive services (IDHS, 2016).

**Benefits of Early Intervention**

There are a multitude of long-term benefits of Early Intervention (Conte, 2018). For example, Early Intervention teaches a child about learning through play and building their social cognitive skills (Conte, 2018). Play improves a child’s brain development and allows the child to successfully overcome challenges in their life through problem-solving, self-control, socialization, and communication (Conte, 2018). Early Intervention also reduces a child’s need for special attention later in life by increasing their academic readiness and social interactions with peers (Conte, 2018). These positive outcomes are found regardless of whether the Early Intervention is parent-implemented or clinician-directed (DeVeney, 2016; Rutherford, 2019). Parent-focused interventions can eliminate stressors in both the parents and the children (Callanan, 2019). Unsurprisingly, parental stress decreases when their children receive proper care for their developmental delay (Callanan, 2019). In short, the years between birth to age three are critical to learn basic, foundational skills of life; Early Interventions targeting these foundational skills have lasting and diffuse benefits. To deliver early, parent-focused interventions, eligibility for Early Intervention services must be accurately determined.

**Assessment Tools for Determining Eligibility**
Norm-referenced (i.e., standardized) tests and criterion-referenced procedures are two common assessment tools used to determine intervention eligibility (IDHS, 2016). Norm-referenced tests provide the clinician with standard scores, standard deviations, and percentile ranks based on a sampled population’s normal distribution (Weiss, 2018). Standard scores incorporate the range of variability in their calculation. One key advantage of norm-referenced tests is that they directly compare a child’s performance to other chronological age-matched children on the same test (Andersson, 2004). This gives the clinician a means to see whether a child’s performance falls below the range of scores typically expected for a child their same age within a broader population sample. The use of standardized scores, rather than AE scores, from norm-referenced tests is preferred for clinical decision-making even though these norm-referenced tests may have some inherent biases depending on the demographics of the children included in the norming sample (Andersson, 2004). For example, the *Receptive One-Word Picture Vocabulary Test* manual states that “When comparing scores of individuals or groups, or comparing performance across tests, standard scores are the preferred index of performance” (Brownell, 2011, p. 35). This is because standard scores incorporate the variability of scores within a normal population distribution (Adeyemi, 2010).

In addition to norm-referenced tests, published criterion-referenced procedures will also provide AE scores (e.g., *The Rossetti Infant-Toddler Language Scale*; Rossetti, 2006). Criterion-referenced procedures, another type of assessment tool, show a child’s developmental functioning concerning specific criteria; they do not compare that child’s performance to that of a chronologically age-matched population. These tools are useful for measuring developmental progress and for establishing intervention targets along a developmental approach (Andersson, 2004). But, criterion-referenced procedures do not compare a child’s performance to their age-
matched peers, nor do they incorporate the normal distribution of scores that would be typical given a child’s age. Because norm-referenced tests and criterion-referenced procedures are designed differently and provide different information, clinicians must be mindful of how they interpret scores when determining intervention eligibility and when reporting assessment outcomes with families.

The Use of Age Equivalent Scores to Determine Early Intervention Eligibility

Because Illinois requires a percent of developmental delay for determining eligibility, both norm-referenced tests and criterion-referenced procedures are frequently used to obtain AE scores to determine eligibility for Early Intervention, even though neither forms of assessment tools are not designed to be used in this way. One caution when using AE scores to determine eligibility for services is that they lack the necessary comparative performance information that could indicate the presence of a disorder or delay (Andersson, 2004). This is why several norm-referenced tests include a limitation section advising the examiner not to use AE scores when determining whether or not a disorder or delay is present nor when determining eligibility for services. For example, although the PLS-5 is recommended for use by the state of Illinois, the same test discourages the use of AE scores for diagnosing a developmental delay or disorder (Zimmerman et al, 2011). The PLS-5 states, “Because it does not give information about the range of scores for children in a specific age group, an age equivalent does not give you the information you need to determine if a child has a language disorder” (Zimmerman et al., 2011, pp. 17). It goes on to state, “An age equivalent may suggest that a child is functioning below his or her chronological age when in reality his or her score is within normal limits for a child that age” (Zimmerman et al., 2011, pp. 18). In other words, because AE scores do not indicate
whether or not a child’s score may be typical for their age, their use can lead to misinterpretations when determining intervention eligibility.

Another limitation of AE scores is that they may be imprecise because they do not sufficiently match the full spectrum of chronological ages (Andersson, 2004). In other words, some chronological ages might not have a corresponding AE score, which could cause a child’s AE score to dramatically shift. A raw score difference of a single point could change the AE score by several months, which is a substantial amount for a child less than three years old. As stated in the PLS-5, “Small raw score changes may result in large changes in age equivalents. Large differences between age equivalents and a child's chronological age may be obtained but interpreting the child's language skills as being far below or far above average for his or her age may be unwarranted because the range of average scores overlaps at adjacent age groups” (Zimmerman et al., 2011, pp. 18). Because of their imprecision, the use of AE scores for determining eligibility for Early Intervention services is worrisome.

An additional limitation of AE scores is that they do not indicate whether a child is performing the same as a child of the given age equivalent (Andersson, 2004). For example, if a 24-month-old child receives a 16-month AE score, it's a misinterpretation to assume the child's performance is the same as that of a 16-month-old. The 24-month-old could have performed better in areas than a 16-month-old or vice versa (Andersson, 2004). Reporting AE scores to families can be misleading—they do not necessarily reflect the age level of their child’s performance. This may unduly exacerbate the emotional strain placed on families if they misinterpret that their child is functioning at a level significantly below their child’s age, which may not be true.
Because of the limitations and potential misinterpretations of AE scores, their use for determining a child’s eligibility for Early Intervention services has not be scientifically supported (Andersson, 2004). Furthermore, the developers of the assessment tools strongly discourage using the derived AE scores for eligibility purposes. As stated in the PLS-5 manual, “Placement decisions or diagnoses should never be based on equivalents only or on any one type of score…You can only make judgments about a child's standing relative to peers by using standard scores or percentile ranks” (Zimmerman et al., 2011, pp. 18). The manuals of several standardized assessments indicate that the use of AE scores can inaccurately represent a child’s performance; these warnings should be the first sign that misclassifications using AE scores when determining eligibility for Early Intervention services could occur in the state of Illinois.

**Misclassifications and Challenges Receiving Early Intervention**

Because of the indisputable benefits of Early Intervention, Part C became a nationwide program to serve children aged birth to 3 who have developmental delays (Rosenberg, et al., 2012). Unfortunately, only a small proportion of children who are eligible for Part C services receive Early Intervention (Rosenberg, et al., 2008). Rosenberg and colleagues studied the percentage of children who were eligible for Early Intervention in each state and compared that to the number of children who received intervention (2012). The proportion of infants and toddlers likely to be eligible for Part C services ranged from 2% to 78% across the United States. However, the proportion of children officially enrolled in Part C only ranged from 1.48% to 6.96% (Rosenburg et al., 2012).

In 2015, about 21,387 children in Illinois received Early Intervention services. In Illinois, there are about 474,588 children under 3 years of age (Poverty Law). This means that about 4% of Illinois children received services for Early Intervention in 2015. Using data from 2010, about
38% of 9-month-old children and 23% of 24-month-old children were eligible for Early Intervention services in Illinois; however only about 4% of children received services (Rosenburg et al, 2012). In Illinois and nationally, many children who are eligible for Part C do not receive the Early Intervention services that they desperately need. Misinterpretations of AE scores for determining Early Intervention eligibility could be contributing to this disparity.

**Research Questions**

The use of AE scores to determine eligibility for Early Intervention services is problematic and not scientifically supported (Andersson, 2004). However, it is unclear to what extent the use of AE scores is impacting eligibility decisions in Illinois. The purpose of this study was to assess the likelihood of misclassifications on a variety of Illinois pre-approved standardized tests assessing early speech and language skills using the Illinois 30% developmental delay eligibility criterion. To address this aim, we asked two research questions.

First, what is the likelihood that a child in Illinois with a developmental delay based on a standard score will not meet the 30% developmental delay using the AE score eligibility criterion? Second, what is the likelihood that a child who scores within the typical range using a standard score will meet the 30% developmental delay eligibility criterion using an AE score? Because of the well-established limitations of AE scores, we predicted that misclassifications that could under- and over-qualify for Early Intervention in the state of Illinois were likely to occur.

**Methods**

**Analyzed Tests**

The tests analyzed in this study were selected from the pre-approved list of tests on the Illinois Department of Human Services' official website (IDHS). From there, the test manuals
were found at the Northern Illinois University (NIU) Speech-Language-Hearing Clinic or the NIU Library. For the IDHS to approve these tests, they must follow strict criteria reviewed by the IDHS-Bureau of Early Intervention for IDHS. The tests must be listed in the Mental Measurement Yearbook Series, be nationally distributed, be age appropriate, be formally justified, and be individually administered. This study analyzes only norm-referenced tests and does not use any criterion-referenced procedures as they do not provide standard scores that incorporate the typical population distribution of scores. The norm-referenced tests analyzed in this study included the Pre-School Language Scale Fifth Edition (PLS-5; Zimmerman et al., 2011), Mullen Scales of Early Learning (MSEL; Mullen, 1995), and Goldman-Fristoe Test of Articulation 2nd Edition (GFTA-2; Goldman & Fristoe, 2000).

**Digitization of Test Manual Data**

Data from test manuals were first converted into a digital form. This process involved compiling pictures of all the test scoring interpretation tables from each test manual. Pictures of all the appendices for ages birth to three were acquired with an iPhone 7 camera. In the PLS-5, tables with the age range 0:0-2:11 from pages 125-139, tables labeled ‘Total Language Standard Scores, Percentile Ranks and Confidence Interval’ from pages 162-163, Table C.1 on page 166 and Table C.2 from page 167 were photographed for digital conversion. In the MSEL, Table C.1 from pages 74-80, Table C.2 from page 83, Table C.3 from page 84 and Table C.4 from page 85 were photographed for digital conversion. Lastly in the GFTA-2, Tables A.1 from pages 72-92, Tables B.1 from pages 228-229 and Table B.2 from pages 230-233 were photographed for digital conversion. The pictures were emailed via Gmail and converted to .jpeg format to be viewed on a computer. The photos were then uploaded to an online Optical Character Recognition (OCR)
website (https://easypdf.com/ocr-online). When the image was submitted to the website, the
digital images were translated into sheets of Microsoft Excel version 16.26 (Microsoft, 2018).

**Digital Conversion Reliability**

The primary investigator visually reviewed the data to ensure that the test manual images
had been correctly translated into Excel. Then, to ensure that all data was digitized accurately, a
second undergraduate coder studying Communicative Disorders at Northern Illinois University
reviewed all of the converted Excel sheets alongside the original manual images, and all
disagreements were resolved through consensus. This process occurred for each test
interpretation table prior to conducting any analyses of the data.

**Preparing Data for Analyses**

To address the research questions, only the table information necessary to calculate
likelihood ratios were kept in the digitally converted tables, including chronological age, raw
scores, standard scores, and age equivalency scores; all other information, such as headings,
confidence intervals, growth scales, extra spaces, etc., were deleted. Before any data was
analyzed, formatting changes and modifications was made to ensure that the data were in the
correct format for code analysis. The PLS-5 test was analyzed using a program written in Visual
Basic Analysis (VBA), the macro-programming language of Excel, and the other tests were
analyzed using Python. The code was switched to Python because VBA couldn’t handle the large
number of computations. Python is a faster programming language. First, all ages were converted
into months. For example, an age representation of 2-10 (2 years and 10 months) would be
converted to 34 months. To reduce the opportunity for human error, this age conversion was
done automatically using an Excel formula and was later checked for accuracy by the primary
investigator. Second, for the program to analyze the data, all ranges of scores or scores with
greater than/less than signs in the Excel sheets had to be modified. If there were < or > symbols, the number entered was changed up or down based on the symbol. For example, >53 was changed to 54 and <100 was changed to 99. Next, any cells that were indicated as Text format were changed to Number format to be read by the program by selecting the “Convert to Number” option for each cell. This ensured that the program could calculate the correct numbers.

In the PLS-5, there were some additional changes needed to format the Excel sheets for the program to read. First, all ranges of scores were changed to the highest value of the range, with the exception of ranges in the last row of a table. For ranges in the last row of a table, both the lowest and highest score were included to define a stopping point for the program (see Figure 1). Figure 1 demonstrates how ranges were converted from their original format to cleaner and readable format for later analysis.

Figure 1:

*Before and After Changes for Range Conversions*

<table>
<thead>
<tr>
<th>Sum of AC and EC Standard Scores</th>
<th>Standard Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100-108</td>
<td>50</td>
</tr>
<tr>
<td>109-110</td>
<td>51</td>
</tr>
<tr>
<td>111-112</td>
<td>52</td>
</tr>
<tr>
<td>113</td>
<td>53</td>
</tr>
<tr>
<td>114-115</td>
<td>54</td>
</tr>
<tr>
<td>116-117</td>
<td>55</td>
</tr>
<tr>
<td>118-119</td>
<td>56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sum of AC and EC Standard Scores</th>
<th>Standard Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td>109</td>
<td>51</td>
</tr>
<tr>
<td>111</td>
<td>52</td>
</tr>
<tr>
<td>113</td>
<td>53</td>
</tr>
<tr>
<td>114</td>
<td>54</td>
</tr>
<tr>
<td>116</td>
<td>55</td>
</tr>
</tbody>
</table>
Second, the AE scores <0 was changed to 0 in all relevant sheets to eliminate the < symbol for analysis. This was done because the presence of symbols and letters automatically formats the cell as a string and not a numeric value. So to confirm accurate data analysis, all symbols were removed. Lastly, ages were all converted into months. For example, the raw score of 20 has an Auditory Comprehension age equivalent score of 1-4 (years-months), so 1-4 was converted to 16 months. When the age ranged from <0-1, it was converted to 1 month.

In the MSEL, there were minor changes that occurred specifically for this test. First, because the focus of this study was on eligibility for children with communication delays, the data for the Gross Motor, Visual Reception, and Fine Motor scales were deleted and only the Receptive and Expressive Language scales were included for later analysis. Second, if there was an <1, it was changed to 0. As mentioned before, this step was done to ensure that the program could read the data (i.e., < and > symbols were not included).

The last test analyzed in this study was the GFTA-2. GFTA-2 is set up with AE scores ranging from 0-80 with a combined Excel worksheet for both male and female data. To make a clear distinction between the two, a second Excel worksheet was created in order to make one section for females and one section for males. From there, data were modified to ensure that no ranges were present by having a Lower AE value and an Upper AE value. Figure 2 shows how ranges were converted to better suit the program. This was done to make sure both the lower and upper value were accounted.

Figure 2: 
*Original vs, Revised Upper and Lower data*
“AE 0-80” Original Data:

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>93-95</td>
</tr>
<tr>
<td>1</td>
<td>93-95</td>
</tr>
</tbody>
</table>

“AE 0-80” Revised Female Data:

<table>
<thead>
<tr>
<th>Raw Score</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>93</td>
<td>95</td>
</tr>
<tr>
<td>1</td>
<td>93</td>
<td>95</td>
</tr>
</tbody>
</table>

**Data Analyses**

Once data were checked for accuracy and all formatting changes were made, they were analyzed using a program written in VBA, or the macro-programming language of Excel, and Python to calculate all of the possible combinations of False Negatives, False Positives, True Negatives, and True Positives. The norm-referenced tests analyzed in this study classify a developmental delay through a normal distribution based on the range of scores acquired from a norm-referenced, population sample. A normal distribution uses the standard deviation to explain the dispersion of all scores from the average score. Using this approach, the standardized test would characterize a score of 100 as the mean, and a score of 85 as one standard deviation below the mean (Andersson, 2004). Because the testing manuals recommend the use of the standard score, rather than the AE score, for diagnostic decision-making, the standard score of 85 (i.e., 1 standard deviation below the mean) was set as the minimum frame of reference for determining the presence of a “true” developmental delay.

The Illinois’ eligibility criterion does not use standard scores that account for this normal population variability, but instead uses a developmental delay of 30% based on an AE score to determine the presence of a delay. As such, the 30% developmental delay was compared to the standard score of 85 for determining the number of True Positives, True Negatives, False...
Positives, and False Negatives on each analyzed test. These classifications were defined as follows:

1. True Positive (i.e., test and IL classify as delayed): Standard score $< 85$ and $\%$ delay $\geq 30\%$
2. True Negative (i.e., test and IL classify as typical): Standard score $\geq 85$ and $\%$ delay $< 30\%$
3. False Positive (i.e., IL classifies as delayed, but test classifies as typical): Standard score $\geq 85$ and $\%$ delay $\geq 30\%$
4. False Negative (i.e., test classifies as delayed and IL classifies as typical): Standard score $< 85$ and $\%$ delay $< 30\%$

For every test, the program analyzed every possible combination of standard score and AE score for all valid chronological ages. Figures 3 and 4 show a general schematic representation of the programming logic that was used to calculate and classify all possible combinations for a test. Each test required its own code due to the uniqueness of each test as explained in the previous discussion of the data cleaning and preparation. For each possible outcome, the percent delay was calculated and compared to the 30% delay that Illinois requires. Importantly, the likelihoods represent the total number of possible combinations of scores rather than indicating the number of cases or children for each classification. Although we do not know how frequently children assessed in Illinois would fall into these combinations, even one combination of scores could lead to potentially numerous misclassifications if it commonly occurs. To clarify, a False Positive score of 1350 would indicate that the program found 1350 combinations of raw scores, chronological ages, AE scores, and standard scores that could lead
to a misclassification of a developmental delay in the state of Illinois, not 1350 cases of children who were misclassified.

Figure 3:

Algorithm 1 Analysis of All Possible Combinations for a Given Test

1: function TestAnalysis(test, chronAge_{\text{min}}, chronAge_{\text{max}}) \quad \triangleright \text{Where test is the Excel file}
2: \quad \text{for } i = chronAge_{\text{min}} \text{ to } chronAge_{\text{max}} \quad \text{do}
3: \quad \text{Find corresponding data sheet with raw scores from } test
4: \quad \text{for each raw score value } r \text{ in sheet } \text{do}
5: \quad \text{Get the } SS \quad \triangleright \text{Where } SS \text{ is the corresponding standard score}
6: \quad \text{Get the } AE \text{ corresponding to } SS \quad \triangleright \text{Where } AE \text{ is the age equivalent}
7: \quad \text{Calculate the } PD \quad \triangleright \text{Where } PD \text{ is the percent delay}
8: \quad \text{Determine } class = \text{CLASSIFY}(PD, SS)
9: \quad \text{Save } class \text{ data}
10: \end \text{ for}
11: \text{end for}
12: \text{end function}

Figure 4:

Algorithm 2 Classification of One Combination

1: function CLASSIFY(PD, SS)
2: \quad class = \text{None}
3: \quad \text{if } PD \geq 30\% \text{ and } SS < 85 \text{ then}
4: \quad class = \text{True Positive}
5: \quad \text{else if } PD < 30\% \text{ and } SS \geq 85 \text{ then}
6: \quad class = \text{True Negative}
7: \quad \text{else if } PD \geq 30\% \text{ and } SS \geq 85 \text{ then}
8: \quad class = \text{False Positive}
9: \quad \text{else}
10: \quad class = \text{False Negative}
11: \text{end if}
12: \text{return } \text{class}
13: \text{end function}

Results

The purpose of this project was to investigate the likelihood of misclassifications based on the Illinois Early Intervention 30% developmental delay criterion. As such, this study aimed to answer two questions: (1) what is the likelihood that a child in Illinois with a developmental delay based on a standard score will not meet the 30% developmental delay using an AE score eligibility criterion (i.e., a False Negative); and (2) what is the likelihood that a child who scores within the typical range using a standard score will meet the 30% developmental delay eligibility criteria using an AE score in Illinois (i.e., a False Positive)? To address these questions, three
assessment tools from Illinois’ pre-approved test list, the PLS-5, MSEL, and GFTA-2, were analyzed using the coded program to identify the number of combinations of scores that could lead to misclassifications. Summaries of the results for the total combinations of for PLS-5, MSEL, and GFTA-2 are presented in Tables 1, 2, and 3 respectively.

Table 1:

**PLS-5 AC and EC Combined Program Output Results**

<table>
<thead>
<tr>
<th>Types</th>
<th>True Positive</th>
<th>True Negative</th>
<th>False Positive</th>
<th>False Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>22720</td>
<td>129689</td>
<td>2320</td>
<td>2351</td>
</tr>
<tr>
<td>Percent</td>
<td>0.144</td>
<td>0.826</td>
<td>0.015</td>
<td>0.015</td>
</tr>
</tbody>
</table>

*Note. AC= Auditory Comprehension; EC=Expressive Communication; VBA= Visual Basic for Applications*

Table 2:

**MSEL RL and EL Subscale Program Output Results**

<table>
<thead>
<tr>
<th>Types</th>
<th>True Positive</th>
<th>True Negative</th>
<th>False Positive</th>
<th>False Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL Number</td>
<td>217</td>
<td>517</td>
<td>6</td>
<td>43</td>
</tr>
<tr>
<td>RL Percent</td>
<td>0.277</td>
<td>0.660</td>
<td>0.008</td>
<td>0.055</td>
</tr>
<tr>
<td>EL Number</td>
<td>180</td>
<td>581</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>EL Percent</td>
<td>0.230</td>
<td>0.742</td>
<td>0.005</td>
<td>0.023</td>
</tr>
</tbody>
</table>

*Note. RL=Receptive Language; EL= Expressive Language*

Table 3:

**GFTA-2 Upper and Lower Program Output Results**

Female:
### Upper Bound Classification

<table>
<thead>
<tr>
<th></th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>TP</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Bound Classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FN</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>84</td>
<td>106</td>
<td></td>
<td></td>
<td>190</td>
</tr>
<tr>
<td>TN</td>
<td>944</td>
<td></td>
<td></td>
<td></td>
<td>944</td>
</tr>
<tr>
<td>TP</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td>838</td>
</tr>
</tbody>
</table>

### Lower Bound Classification

<table>
<thead>
<tr>
<th></th>
<th>FN</th>
<th>FP</th>
<th>TN</th>
<th>TP</th>
<th>Grand Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FN</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP</td>
<td>135</td>
<td>92</td>
<td></td>
<td></td>
<td>227</td>
</tr>
<tr>
<td>TN</td>
<td>981</td>
<td></td>
<td></td>
<td></td>
<td>981</td>
</tr>
<tr>
<td>TP</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>775</td>
</tr>
</tbody>
</table>

**Note.** FN=FALSE NEGATIVE; FP=FALSE POSITIVE; TN=TRUE NEGATIVE; TP=TRUE POSITIVE

### The Likelihood of False Negatives When Using AE Scores

On the PLS-5, the program uncovered 2,351 possible score combinations that could lead to a False Negative classification, or a standard score that would indicate the child’s performance is more than one standard deviation below the typical range but would not meet the Illinois 30% developmental delay criterion. On the MSEL, the program identified 43 possible score combinations on the Receptive Language Subscale and 18 possible score combinations on the Expressive Language Subscale that could lead to a False Negative classification.

The GFTA-2 was separated in two categories: Female and Male (see Table 3). From there, each category had two subcategories of lower and Upper Bound. For the female Lower Bound, there were 10 False Negative combinations; the Female Upper Bound had 16 False Negative score combinations. The male Lower Bound had 4 False Negative combinations and 5
False Negative combinations for the Upper Bound. For example, a 32-month-old female with a raw score of 80 would obtain a standard score of 82, a Lower Bound AE score of 0, and an Upper Bound AE score of 23. The percent developmental delay calculated with the Lower Bound would be 100%, which was considered a True Positive, and a percent developmental delay of 28.1% with the Upper Bound, which was considered a False Negative. Overall, the GFTA-2 had a combined, both female and male, classification of 14 False Negative combinations in the Lower Bound and 21 combinations in the Upper Bound.

In total, 2,447 False Negative combinations of scores were identified across all three of the tests analyzed in this study. Importantly, the program uncovered False Negative combinations that could lead to the denial of services for children who are performing below the typical range on every single one of the Illinois pre-approved assessment tools analyzed in this study.

**The Likelihood of False Positives When Using AE Scores**

On the PLS-5, the program revealed 2,320 possible score combinations that could lead to a False Positive classification, or a standard score that classified the child’s performance as delayed while the test classifies the child’s score as typical. On the MSEL, the program identified 6 possible score combinations that could lead to a False Positive, or an over diagnosis, on the Receptive Language Subscale and 4 possible score combinations on the Expressive Language Subscale.

When looking at the GFTA-2 results specifically (see Table 3), there were combinations where a conflict occurred between the Lower and Upper Bound in the classification. In the GFTA-2, the female Lower Bound had 190 False Positive combinations and the Upper Bound had 84 False Positive combinations. The Male Lower Bound had 227 False Positive
combinations and the Upper Bound had 135 False Positive combinations. For females, of the 190 False Positive classifications by the Lower Bound, 106 (55%) of the combinations were classified as True Negative when using the Upper Bound. For males, there were 227 combinations for a False Positive in the Lower Bound. However, 92 (40%) of the same combinations were classified as True Negative when using the Upper Bound. For example, a 29-month-old male on the GFTA-2 with a raw score of 71 would obtain a standard score of 95, a Lower Bound AE score of 0, and an Upper Bound AE score of 23. Using the Lower Bound, this child would receive a percent developmental delay of 100%, which was considered a False Positive, whereas using the Upper Bound would lead to a percent developmental delay of 20.6%, which was considered a True Negative. For the combined male and female results on the GFTA-2, there were 417 False Positive combinations using the Lower Bound and 219 combinations using the Upper Bound. These conflicting outcomes indicate that, depending on which AE score within the provided range was selected, the child’s same raw score could lead to a nearly random chance of being correctly classified. These conflicts also indicated that Upper Bound results showed better alignment between the standard score and the Illinois developmental delay criterion when determining eligibility for Early Intervention.

In total, 2,966 False Positive combinations of scores were identified across all three of the tests analyzed in this study. Similar to the False Negative results, the program uncovered score combinations on all three of the Illinois pre-approved assessment tools that could cause children who performed within the typical range based on their chronological age to be eligible to receive Early Intervention services in Illinois.

Discussion
As predicted, the use of AE scores to determine eligibility for services in Illinois is problematic, as shown by the thousands of combinations of scores uncovered in this study that could lead to misclassifications on these three tests alone. GFTA-2 results indicate that a child is more likely to be over diagnosed, rather than under diagnosed. These results are opposite from what occurred in the PLS-5 and MSEL. The PLS and the MSEL, there were more False Negatives than False Positives most likely results in a denial of early intervention services, whereas the GFTA had more False Positives rather than False Negatives. Consistent with the recommendations of previous research (e.g., Andersson, 2004) and as explicitly stated in the test manuals (Zimmerman et al., 2011; Mullen, 1995; Goldman & Fristoe, 2000), the results of this study indicated that the use of AE scores to determine eligibility for services should be strongly discouraged. In fact, on all tests, combinations that could lead to the denial of services for children with developmental delays (i.e., False Negatives) and the inappropriate provision of services to children without a true developmental delay (i.e., False Positives) were identified.

A denial of necessary Early Intervention services could have long-term consequences for these children. Cognitively, children could fall behind on information processing and decision-making skills (Cherry, 2020). Furthermore, these cognitive skills affect a child’s social and play skills as they are necessary for a child to successfully communicate with their peers. If a child is misclassified, they could be denied services tailored to improve their communication and social interactions. As the child goes on and enters school settings, a child’s interactions play a role in their personal and academic life, causing them to fall further behind in school (Cherry, 2020). In addition to the lifelong impact of false classifications on children, these misclassifications also could lead to an increase in parental stress. When children are denied necessary services, parents are left to advocate to acquire the best resources for their child to be successful. Overall, the
misuse of AE scores to determine Early Intervention eligibility does not help children or families in the state of Illinois.

Although the False Negatives are especially concerning because of the lifelong consequences of preventing children from receiving Early Intervention, the False Positives identified in our results also have devastating consequences for families and taxpayers. Based on the PLS-5, MSEL, and GFTA-2 results, thousands of False Positive combinations could cause an early interventionist to over-qualify children for services who do not need them. Over-qualification leads to wasted tax dollars for services to children who do not have a true developmental delay. In a way, these misclassifications take away services from a child who needs the resources. Furthermore, False Positives could cause families who go through this process to end up wasting their time and their money paying for intervention services their child does not need. Finally, and importantly, misclassifying children as having a developmental delay when one does not exist also places the family under needless emotional stress.

Using previously reported national data, Rosenberg and colleagues (2012) investigated what percent of children would qualify for Early Intervention using each state’s respective eligibility criteria. These investigators discovered that “The criteria of 18 states categorize 40% of 9-months-olds as delayed. Criteria that make 40% of the population candidates for Part C create serious difficulties by classifying children who fall within the average range of abilities as developmentally delayed. It seems clear that states using the broadest definitions of delay could mischaracterize large numbers of children as delayed and make far too many children candidates for Part C services” (Rosenberg et al., 2012, p. 41). To ensure that resources are distributed efficiently and effectively, it’s essential that eligibility criteria accurately identify those who truly need services.
Determining Eligibility

Data from the PLS-5, MSEL, and GFTA-2 suggest that the use of AE scores leads to thousands of possible combinations of scores that provide both False Positive and False Negative classifications. These results are consistent with the warnings made by previous researchers about the potential for AE scores to lead to false classifications. As stated by Maloney and Larrivee, “AE scores alone can lead to gross misinterpretations of children’s skills and, therefore, are less reliable than other types of scores….The use of standard scores would also allow researchers to compare children’s scores across language modalities and across related skills, such as cognition” (2007, p. 88). It has been thoroughly, empirically established by numerous researchers that AE scores can lead to an inaccurate representation of a child’s development. Although early interventionists may be familiar with the well-researched limitations of AE scores and are trained to implement evidence-based practice, they are mandated by the state of Illinois to make eligibility decisions counter to the best available scientific evidence.

To better understand how the Illinois eligibility criteria were initially established, the principal investigator sought the assistance of legal and governmental library specialists, but she was still unable to locate how the Illinois eligibility criteria were established as official policy. This information would be helpful for researchers and practicing clinicians to be aware of how eligibility criteria are determined and what factors contribute to the policy decision-making process. When looking at other states, many states consider standard scores alongside AE scores to determine eligibility. For example, in Indiana, the criteria are a 25% developmental delay or a standard score of at least 2 standard deviations below the mean in one or more developmental domains (States’ and territories’ definitions of/criteria for IDEA Part C eligibility, 2015). Allowing for the use of standard scores and AE scores could provide clinicians more flexibility
to rely on their expertise during clinical decision-making. If Illinois was to adapt to a system that incorporated standard scores, empirical evidence, and expert clinical judgment, it could decrease the number of children who are misclassified. Illinois could change its eligibility criteria to better reflect the best available scientific evidence and to be more in-line with other states that use standard scores or clinical expertise rather than arbitrary, and misleading, AE scores.

Consistent with previous research, the use of AE Scores in this study was found to introduce opportunities for classification errors. As has been discussed by other researchers and in several of the analyzed tests’ administration manuals, the test developers warned against the use of AE scores for determining intervention eligibility. It would be prudent for the policymakers of Illinois to follow the guidance put forth by Andersson, which explains that if an examiner uses scores in a manner that is not specifically recommended by the developer of the test, it is the examiner’s responsibility to find evidence of the validity of that use of test scores (2004, p. 59). This burden should be on Illinois policymakers to provide evidence of the validity of using AE scores to determine intervention eligibility. This mandated misuse of AE scores puts practicing clinicians in a difficult position; they must either adhere to the strict guidelines for interpreting the test scores provided in the administration manuals by the test developers and in accordance with best evidence or use the scores in the manner mandated by state of Illinois.

**Clinical Implications**

To prevent misinterpretations of AE scores, clinicians need to advocate for their clients based on standard scores and their expert clinical judgment. More autonomy should be given to clinicians who have expertise and training to make diagnostic decisions rather than mandate that eligibility be determined by misleading data. It is a massive financial burden for families to unnecessarily pay for services for their child who is falsely identified with a developmental
delay. Furthermore, parents should not be forced to exert additional energy, time, and financial resources advocating for their child perpetually and falsely be denied the services that they need. For this reason, clinicians need to be aware of how standard and AE scores are derived and how to accurately interpret results to better ensure that families receive the proper services. Illinois clinicians must advocate for their right to exercise good clinical decision-making and to change the Early Intervention eligibility criteria to reflect the best, current scientific evidence.

The American Speech-Language Hearing Association (ASHA) recommends using the best available evidence when assessing children for developmental delays (ASHA, n.d.). Evidence-based practice facilitates positive outcomes for children and their families (ASHA, n.d.). The use of AE scores to determine intervention eligibility is not supported by evidence, so by mandating their use, Illinois is forcing clinicians into ethically precarious positions by having to choose between adhering to best evidence as instructed by their national licensing body and following State policy. Clinicians should be allowed to use their expertise to score and interpret assessment tools as instructed within the administration manuals and following the best scientific evidence when assessing and providing care for children in Illinois.

Limitations

Although the results of this study provide further evidence of the limitations associated with AE scores, this study itself contains limitations of its own. First, because of the need to compare AE scores to standard scores as a frame of reference, this study could not include one of the most widely used assessment tools that provides AE scores—criterion-referenced procedures. Although criterion-referenced procedures are commonly used in assessment, they do not provide an accurate way to determine if a child’s development is typical for their chronological age (Rosenberg & Zhang, 2010). Because Illinois allows for the use of criterion-referenced
procedures to acquire AE scores, the misuse of AE scores is likely more far-reaching than this study can determine. On the three tests included in this study, thousands of misleading combinations of scores were identified, indicating that the inappropriate use of AE scores derived from criterion-referenced procedures is potentially leading to massive misclassifications.

Future Directions

We plan to advocate and spread awareness to Illinois families, clinicians, and policymakers about potential misclassifications when using AE scores to determine Early Intervention eligibility. Based on these results, one recommendation for policymakers would be to address and revise the eligibility criteria to use standard scores and clinical expertise rather than AE scores (Rosenberg and Zhang, 2010). The use of AE scores should be eliminated. The second recommendation would be to track and monitor if children who are eligible for services actually receive those services in accordance with Part C of IDEA (Rosenberg and Zhang, 2010). We hope that the state of Illinois will remove the unsubstantiated and invalid 30% developmental delay criterion and instead move to eligibility criteria based on standard scores or percentile ranks that reflect the variability in the population along with trusting expert clinical judgment.

Conclusions

Overall, this project aimed to determine the likelihood of misclassifications on several pre-approved standardized tests assessing early speech and language skills using the Illinois Early Intervention eligibility criteria. Through this research, we were able to discover that the use of AE scores could lead to misclassifications in many score combinations. The simple use of standard scores and expert clinical judgment could help to prevent misclassifications from occurring and more effectively provide children access to the services they need.

Acknowledgments
I would like to thank Dr. Christine Nguyen from the Northern Illinois University College of Engineering for creating the program that provided results for this project. Initial data collection was supported by Northern Illinois University’s Summer Research Opportunity Program (SROP). The current research analyses were supported by the Center for the Interdisciplinary Study of Language and Literacy (CISLL) at Northern Illinois University. I would also like to extend my gratitude to Onna Hills for her help on this project as a reliability checker. Finally, I would like to thank all the contributors and their families for making this research possible.


https://doi.org/10.1007/s10995-012-0982-2


