ABSTRACT: The literature for speech-language pathology is replete with criticisms of age-equivalent (AE) scores as descriptions of performance on standardized tests. However, there are few empirical examples that demonstrate explicitly the confusion that results from AE scores. This study investigated scores for 19-year-old participants on the Peabody Picture Vocabulary Test—III (PPVT–III; L. M. Dunn & E. M. Dunn, 1997) and examined the relationship between the participants’ standard scores and their AE scores. Although the range of normal for standard scores was 88–115, derived AE scores for these participants ranged from 14;0 (years;months) to 22+. Results revealed that, as expected, standard scores were normally distributed. AE scores were unevenly distributed and skewed, making interpretation problematic. This information can be used to assist speech-language pathologists in implementing evidence-based practice when reporting performance on standardized tests.

KEY WORDS: age-equivalent scores, norm-referenced tests, standardized tests, measurement

Limitations of Age-Equivalent Scores in Reporting the Results of Norm-Referenced Tests

Emily S. Maloney
Linda S. Larrivee
Worcester State College, Worcester, MA

When evaluating and diagnosing clients with suspected language impairment, speech-language pathologists (SLPs) rely on a variety of sources to aid in assessment. These sources include information from parents and teachers, formal observation, and language sample analyses. To obtain objective data on a client’s language abilities, SLPs often rely on standardized tests. The scores on these tests are used to determine whether a client is performing within the average range or whether a delay or impairment is present.

One of the more common types of scores available for reporting performance on standardized tests is age-equivalent (AE) scores. AE scores represent the mean or median score derived for a normative sample for a particular age group. For example, an AE score of 5;0 (years;months) corresponds to the average raw score obtained by 5-year-olds in the normative sample. Most often, AE scores are used to assist SLPs and other professionals in explaining a client’s performance on a standardized test to parents and teachers. In addition, these types of scores are often mandated by agencies at the local, state, and federal level for eligibility and funding purposes (Lawrence, 1992). Despite the popularity of these scores for explanation and funding purposes, AE scores have serious, well-documented limitations (Aiken, 2003; Anastasi & Urbina, 1997; Couzens, Cuskelly, & Jobling, 2004; Kerr, Guildford, & Bird, 2003; Lawrence, 1992; McCauley & Swisher, 1984; Salvia, Ysseldyke, & Bolt, 2006).

Norm-Referenced Standardized Tests

Most standardized tests that are used by SLPs are norm referenced. The developers of norm-referenced tests
determine normative standards that indicate what the
average scores are for relevant age groups by administering
the tests to a representative sample group (Shipley &
McAfee, 2004). The performance of the normative group is
used to develop the normal distribution that serves as a
standard of comparison with which test takers can be
compared. Norm-referenced tests are invaluable tools in the
field of speech-language pathology. They work in conjunc-
tion with other sources of assessment information to
provide objective data to support the presence or absence of impairment.

Types of Scores

AE scores are just one of a number of types of scores that
can be calculated for norm-referenced tests to help an
examiner analyze a test taker’s performance. Most of these
types of scores are determined by first calculating the test
taker’s raw score. A raw score represents the total number of
stimulus items that the test taker answers correctly. Raw
scores are insignificant in interpreting performance because
there is no basis for comparison to other test takers’
performances. To gain perspective on a client’s performance
on a norm-referenced test, the examiner must determine
how the client performed in relation to other individuals of
the same age.

One of the most common types of scores used for
reporting performance on standardized tests is a standard
code. A standard score, which is derived from a raw score,
indicates how many standard deviations away the test taker’s
score is from the mean for a particular age group. These
scores take into consideration the mean score and the
variability of scores within the normative sample. Therefore,
these scores report a range of normal to which a test taker’s
scores can be compared. Standard scores follow an equal-
interval scale and can be used to denote progress over time.
Standard scores can also be used to compare clients’ scores
across language modalities or with performances on tests
evaluating related areas, including cognition and reading.

Percentile ranks are another type of measure that can be
used to report performance on standardized tests. Percentile
ranks indicate an individual’s standing within the normative
sample. A test taker’s percentile rank on a given test
indicates the percentage of examinees in the normative
sample that scored above and below the test taker on the
test. Percentile ranks are useful because they allow an
individual’s score to be compared to the normative sample.
These types of scores also describe a range of normal in
which an individual’s score can compared. Nonetheless,
care must be taken when interpreting percentile ranks. As
percentile ranks near the high and low ends of the norma-
tive sample, small discrepancies between percentile ranks
can be reflective of larger differences in raw scores.

Limitations Associated With AE Scores

As mentioned, AE scores are a commonly used method of
reporting test performance. AE scores assist SLPs when
explaining test takers’ performance to parents and teachers.
In addition, they have been used consistently by local,
state, and federal agencies to establish eligibility and
funding guidelines for children. Despite the wide use of AE
scores, there are several well-documented limitations
associated with these scores. First, in contrast to standard
scores and percentile ranks, AE scores do not take into
consideration the range of normal performance for individu-
als whose scores fall within the average range. Rather,
these scores represent the age at which a given raw score
is average (McCauley & Swisher, 1984). It would be
expected that half of the examinees on a test will achieve a
higher AE score than their corresponding chronological age
(CA) (Lawrence, 1992). Similarly, half of the examinees
should receive a lower than average AE score. Salvia et al.
(2006) pointed out that the lack of consideration for a
range of normal performance results in AE scores implying
a false standard of performance. For example, one might
expect a 4-year-old child to earn an AE score of 4.0.
However, due to the nature of AE scores, half of the 4-
year-old examinees will earn an AE score that is below
their CA. A child who receives a standard score or percen-
tile rank that is below the mean for a given age group may
be performing well within the range of normal, or within 1
SD away from the mean. This same examinee might earn an
AE score that is significantly below his CA. Therefore,
AE scores make no attempt to describe a normal range of
performance, making these types of scores ineffective in
making case management decisions (Lawrence, 1992).

A second reported limitation of AE scores is that these
scores promote typological thinking (Salvia et al., 2006).
AE scores compare children to the “average x-year-old.”
However, the average x-year-old does not exist. Rather,
the term average represents a range of performance for a
particular age group.

A third serious limitation of AE scores is the lack of
information they provide about a test taker’s performance
on a given test. According to Lawrence (1992), when two
children earn the same AE score, the examiner cannot
assume that the children responded the same way to the
stimulus items on the test. Earning the same AE score
simply means that these two children answered the same
number of questions correctly. Although a 5-year-old and a
10-year-old may earn the same AE score, these two
children may have approached the stimulus items differ-
ently. That is, they may have demonstrated varying
performance patterns. It is likely that the younger child
performed lower level work with greater consistency,
reaching a ceiling early on. The older child likely at-
tempted more problems but performed at a lower accuracy
level (Lawrence, 1992; Salvia et al., 2006). Consequently,
AE scores would be ineffective in making inferences about
what can be expected from these children regarding their
language abilities (McCauley & Swisher, 1984).

AE scores also may be ineffective in assessing children
with severe developmental delays or mental retardation.
Couzens et al. (2004) suggested that AE scores are not
valid when evaluating children with Down syndrome
because these children may use different underlying
processes when approaching stimulus items. The authors
examined the difficulties that occur when using the
Stanford Binet—Fourth Edition (Thorndike, Hagen, &
Despite the well-documented limitations of AE scores, these months for a 14-year-old. Therefore, AE scores cannot be differences in raw scores” (McCauley & Swisher, 1984, p. equivalent scores are the result of smaller and smaller flattening of the curve as age increases (Aiken, 2003; follow an equal-interval scale, AE scales are ordinal, with a year to year (Aiken, 2003). Unlike standard scores, which falsely imply that abilities increase at a constant rate from 5:0 and 5:5. These children’s scores are plotted and smoothed into a graph. Using this graph, the AE scores for children at each month interval are estimated, or extrapolated. The AE score for each age represents the average raw score for that age group of children (Lawrence, 1992). Thus, when AE scores are calculated, they represent a mean score of a group of children who were not actually tested (Salvia et al., 2006). A fifth problem relates to children who receive extremely high or low scores. For these children, the AE scores are reported as higher or lower than a given number, respectively. A child who performs particularly poorly on a test may receive an extremely low raw score for which an AE score was not extrapolated. For example, a 10-year-old child who scores very poorly may receive an AE score such as <4:0 because AE scores were not derived for CAs below 4 years. In these cases, AE scores offer limited information to the examiner regarding the child’s performance on the test. A sixth problem with AE scores is that these scores falsely imply that abilities increase at a constant rate from year to year (Aiken, 2003). Unlike standard scores, which follow an equal-interval scale, AE scales are ordinal, with a flattening of the curve as age increases (Aiken, 2003; Anastasi & Urbina, 1997; Lawrence, 1992; Salvia et al., 2006). That is, “as age increases, similar differences in age equivalent scores are the result of smaller and smaller differences in raw scores” (McCausley & Swisher, 1984, p. 34). For example, a difference in AE scores of 3 months for a 4-year-old is more significant than a difference of 3 months for a 14-year-old. Therefore, AE scores cannot be used to demonstrate change in a child’s skills over time (Lawrence, 1992).

Use and Prevalence of AE Scores

Despite the well-documented limitations of AE scores, these types of scores are often used in speech and language evaluations to diagnose impairment. McCauley and Demetras (1990) examined articles reporting research designs involving children with specific language impairment (SLI) and found that test data were the most frequently used source for diagnosing children with SLI, and AE scores were the most commonly reported type of score. More specifically, in 45% of the articles examined, AE scores were the only score reported. In an additional 17% of the articles, AE scores were reported in conjunction with other types of scores. In contrast, scores that include information regarding the variability of the normative sample, such as standard scores, were only reported in 24% of the articles. McCauley and Demetras concluded that this study exemplifies the overreliance on AE scores to identify impairment—a common misuse of these types of scores. Because these scores do not take into consideration individual differences within the normal range, diagnosing language impairment based on AE scores alone can lead to gross misinterpretations of children’s skills and, therefore, are less reliable than other types of scores. In response to their findings, McCauley and Demetras suggested that researchers of SLI should consider the benefits of using standard scores, which allow them to operationalize definitions of language impairment through the use of cutoff scores and avoid using AE 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. The common use of AE scores in speech and language evaluations also was documented by Kerr et al. (2003). These researchers examined standardized test use by a sample of SLPs practicing in Canada to identify misuses associated with standardized tests. Kerr et al. found that 6% of the SLPs reported always using AE scores to summarize test results, 21% reported frequently using these types of scores, and 47% reported using them sometimes. In contrast, only 26% of the SLPs stated that they never use AE scores when summarizing test results. Seventy-six percent of the SLPs reported that there were benefits to using these types of scores. Reasons listed by respondents in this survey for using AE scores included that these types of scores are useful when providing information to teachers or parents, to secure funding, to report scores when norms do not apply, and as a gross or initial measure to evaluate the severity of disorders and to guide additional testing. Although these are appropriate uses of AE scores, other responses indicated that many SLPs misunderstood the inherent limitations of these scores. This misunderstanding could result in conveying misleading information to parents or teachers. For example, some SLPs reported that AE scores are useful for comparing children to their peers and for measuring treatment progress. However, these two uses of AE scores can lead to potential misuse and misinterpretation. Although many SLPs who were surveyed by Kerr et al. (2003) reported the benefits of using AE scores, they also reported limitations of these types of scores. One commonly reported problem was that a large AE delay may not be indicative of impairment because a large delay is common in normal children in a given age group. The SLPs also noted that an AE score that is significantly
below a child's CA does not mean that the child has the language skills or world knowledge denoted by that score. Additional problems reported included the lack of usefulness of AE scores when explaining test performance to parents and teachers due to the fact that these scores may sound overly threatening, evoke negative labeling, and lead to low expectations. They also are not helpful in establishing goals or measuring progress in treatment (Kerr et al., 2003).

It is evident that AE scores have significant limitations and can lead to misuse and misinterpretation of test performance. The American Psychological Association has advocated that these scores no longer be used (as cited in Carrow-Woolfolk, 1999). Yet, these scores continue to be used voluntarily by many professionals and are still mandated by many agencies at the local, state, and federal level to establish eligibility and funding for services. Most standardized tests continue to report AE scores, although many test authors explain the limitations of these scores and caution against their use in the examiner's manual (Carrow-Woolfolk, 1999; Semel, Wiig, & Secord, 2003).

Despite the well-documented limitations associated with AE scores, it appears that these scores will continue to be used to report performance on standardized tests. If this is the case, additional research is needed to increase SLPs' understanding of these types of scores, including the ways in which they may be misused. With an increasing focus on using evidence-based practice, SLPs must integrate current, high-quality research evidence with clinical expertise and client preference when making decisions regarding evaluation and treatment. Specifically, SLPs and audiologists should "evaluate prevention, screening, and diagnostic procedures, protocols, and measures to identify maximally informative and cost-effective diagnostic and screening tools, using recognized appraisal described in the evidence-based practice literature" (American Speech-Language-Hearing Association, 2005). When reporting performance on standardized tests, SLPs must refer to research evidence to determine which standardized test scores are maximally informative and which lead to misinterpretation and misuse.

The present study examined how AE scores described performance on the Peabody Picture Vocabulary Test—III (PPVT–III; Dunn & Dunn, 1997) for a group of 19-year-olds whose scores fall within the range of normal, as determined by standard scores. The results of this study may provide empirical evidence for theoretical limitations associated with AE scores. Thus, this evidence can be used to assist SLPs in implementing evidence-based practices when reporting performance on standardized tests.

**METHOD**

**Participants**

Fifty individuals ranging in age from 19:0 to 19:11 participated in this study. Based on analysis of the normative information in the PPVT–III, 19-year-olds were selected as the optimal age for two reasons. First, for examinees ages 18:11 and younger, standard scores are derived for age groups spanning 2–3-month increments. However, after the age of 18:11, the next age range includes ages 19:0 through 20:11. Therefore, the standard scores for all 19-year-olds are calculated using the same norms table. This uniformity makes it possible to compare raw scores, standard scores, and AE scores for all 19-year-olds. Second, extrapolation of AE scores from raw scores tends to be more sensitively defined for lower raw scores. That is, for raw scores on the lower end of the continuum, AE scores increase in 1-month increments. In contrast, as raw scores increase, the corresponding AE scores become decreasingly sensitive. For example, a raw score of 171 corresponds to an AE score of 19:1, and a raw score of 172 corresponds to an AE score of 19:9. Thus, an increase of 1 in the raw score reflects an 8-month increase in the AE score. Therefore, very minimal changes in raw scores result in much larger discrepancies in AE scores. These large discrepancies as the result of minimal increases in raw scores are in contrast to the much smaller discrepancies of AE scores for the same increases in raw scores at the lower end of the spectrum. For these reasons, 19-year-olds appeared to be an age group whose AE scores on the PPVT–III were particularly susceptible to misinterpretation and misuse. Consequently, this age group was determined to be an optimal starting point for research attempting to provide evidence for the limitations of AE scores.

Thirty-two percent of the participants were male and 68% were female. Ninety-eight percent of the participants were college students, with 34% being freshman and 34% being sophomores. The remaining participant was not currently enrolled in college at the time of the study but had plans to enroll in college in the next semester. The participants who were college students indicated a variety of majors/academic areas of interest (see Table 1). All of the participants were fluent in English and had normal hearing. Ninety-four percent of the participants were native English speakers. Additional native languages spoken by the participants included Punjabi, Hindi, Arabic, and Vietnamese. These participants indicated that they were fluent in English as well as in their native language. In addition, all students whose native language was not English had passed an English proficiency test as part of their college entrance requirements. Four percent of the participants indicated that they had some type of learning disability (language, attention deficit disorder). The remaining participants indicated that they were unaware of the presence of any type of learning disability.

**Procedure**

Each participant attended one testing session. The sessions took place in a quiet room, with only the participant and the examiner present. Each participant was presented with a one-page questionnaire with questions pertaining to the participant's gender, birth date, major/academic area of interest, year in school, native language, presence of a learning disability, and hearing status. On completion of the questionnaire, the examiner administered the PPVT–III Form A according to the administration guidelines outlined in the examiner’s manual. The PPVT–III protocol was scored after the participant left the room. The majority of
testing was conducted by the first author. The remaining
testing was completed by first-year graduate students in a
master’s degree program in speech-language pathology who
had been trained in administration of the PPVT–III.

After all 50 of the 19-year-olds had participated in the
study, the PPVT–III protocols were compiled and analyzed.
Because the purpose of this study was to examine the
effectiveness of AE scores in reporting performance on a
standardized test for a group of individuals performing
within the average range, all participants whose scores fell
outside the range of normal were excluded from the
analysis. The range of normal was determined to be those
standard scores that fell within 1 SD from the mean (M =
100, SD = 15). Therefore, only those participants whose
standard scores were between 85 and 115 were included in
the analysis. As a result, 7 of the participants were
excluded from the study. All 7 of these participants earned
scores that exceeded the range of normal. The scores for
the remaining 43 participants were included for analysis.

RESULTS

The mean CA for the participants was 19:6. The CAs
ranged from 19:0 to 19:11. Frequency distributions were
plotted for raw scores, standard scores, and AE scores. The
following results were obtained for the participants scoring
within the average range (n = 43).

Figure 1 shows the participants’ raw scores. The raw
scores ranged from 155 to 186 and had a mean of 174.4.
The raw score plot was asymmetrical, skewed positively.
The majority of participants scored within the 182–186
range. This figure indicates that although the range of raw
scores was large, the majority of the participants earned
raw scores that were at the higher end of the range.

The participants’ standard scores are represented in
Figure 2. The standard scores ranged from 88 to 115 and
had a mean of 103.9. The standard score plot generally
approached the normal curve. This plot demonstrates that
when standard scores are used to report performance for
this group of 19-year-olds, the participants’ scores fall
within the average range and reflect a distribution that is
typical of the normal population.

The range of AE scores was 14:0 to 22+ and is shown in
Figure 3. The AE plot was skewed toward the upper end of
the possible AE scores. As can be seen in Figure 3, the
scores were fairly evenly distributed across the graph until
a sharp increase of scores occurred as the plot approached
22+. This figure indicates that the majority of participants
(n = 24) earned AE scores of 22+, with much fewer
earning scores at each of the potential lower AE scores.

The difference between each participant’s CA and AE
scores was calculated. Figure 4 shows the frequency
distribution of these differences in scores. Although all
participants had standard scores that were within normal
limits for their CA, 12 of the participants (30%) earned AE
scores that were below their CA. Of these 12 participants,
8 earned an AE score that was 2 or more years below their

Table 1. Majors/academic areas of interest (N = 49).

<table>
<thead>
<tr>
<th>Major/academic area of interest</th>
<th># participants</th>
<th>% participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Business</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Communication disorders</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Nursing</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Communications</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Undeclared</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Biotechnology</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Education</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Biology</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Sociology</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Psychology</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Math</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Computer science</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>English</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Urban studies</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>English literature and Spanish</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Landscape contracting</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Technical communications</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Political science</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Chemical engineering</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Exercise science</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
CA. The remaining 31 participants (72%) earned AE scores that were greater than their CAs. Of these 31 participants, 24 earned AE scores that were more than 2 years above their CAs. Thus, out of the 43 participants scoring within the average range, 32 (74%) earned AE scores that were 2 years above or below their CAs.

**DISCUSSION**

**Limitations of AE Scores**

In general, the results of this study support the concerns proposed in the literature associated with AE scores. Out of the 50 participants included in this study, 43 performed within the average range when average was considered to be within 1 SD from the mean. Although these standard scores fell within 1 SD from the mean, the range of AE scores for these same participants was wide. The participants earned AE scores that ranged from 14;0 to 22+. This range suggests that a 19-year-old who earns a standard score of 88, which is well within the range of normal based on standard scores, could earn an AE score of 14;0, a score that is more than 5 years below his CA. The results from this study provide evidence for the potential for AE scores to be grossly misinterpreted by individuals who are unfamiliar with the nature of AE scores. Although professionals administering standardized tests may understand that an examinee earning a standard score of 88 and an AE score of 14;0 is functioning in the normal range, these results may not be understood accurately when being explained to a parent or teacher. It is likely that parents may be concerned to hear that their child earned an AE score that is 5 years below his or her CA.

The results of the study also provide evidence for the concern that AE scores promote typological thinking. For example, consider the examinee who earned an AE score of 14;0. This AE score implies that the examinee earned a raw score that was average for a group of 14-year-olds, or in other words, performed similarly to the average 14-year-old. However, as Salvia et al. (2006) pointed out, the average 14-year-old does not exist. Rather, the term average implies a range of performance for a particular age group. Moreover, even if this examinee answered the same number of questions correctly as the average 14-year-old, this score offers limited information about his or her performance on the test. Lawrence (1992) explained that even though a 19-year-old and a 14-year-old may have earned the same AE score, the examiner cannot assume that the two examinees responded in the same way to the stimulus items. For example, the 19-year-old may have attempted more problems but performed with a lower accuracy level, whereas the 14-year-old may have performed more consistently but reached a lower ceiling. These two examinees’ identical AE scores simply mean that by the end of the test, they answered the same number of items correctly.

The findings from this study also provide evidence for the limitation that is posed for examinees who earn extremely high or low scores. Many of the participants in this study earned AE scores of 22+. In fact, 24 of the 43 participants (56%) earned AE scores of 22+. Participants who earned standard scores between 104 and 115 all earned AE scores of 22+. Because AE scores are not derived for examinees beyond a given raw score, all participants who receive raw scores beyond 177 earn an AE score of 22+. Therefore, AE scores offer limited information regarding these examinees’ performance on the test. When using AE scores, it appears that an examinee earning a standard score of 104 and an examinee earning a standard score of 115 performed the same on the test. That is, they both earned an AE score of 22+. Clearly, the difference between a standard score of 104 and a standard score of 115 indicates a drastic difference in performance. The use of percentile ranks demonstrates this difference. A standard score of 104 corresponds to a percentile rank of 61, whereas a standard score of 115 corresponds to a percentile rank of 84. This finding provides strong evidence of how standard scores and percentile ranks are more effective in reporting performance than AE scores, particularly for examinees earning very high scores.
The limitations associated with AE scores applied to all but a small percentage of the participants. Thirty-two of the 43 participants earned AE scores that differed from their CA by at least 2 years. Yet, all of these participants performed within the average range. It seems that a 2-year separation between AE scores and CAs is likely to lead to misinterpretation of scores. Therefore, it can be concluded that, for this study, AE scores were ineffective in reporting performance for 74% of the sample.

Another limitation that is worthy of note but that may not have been directly evidenced in this study is the method in which AE scores are derived. As previously mentioned, AE scores are derived through interpolation and extrapolation. That is, test makers plot and smooth a graph of scores for a particular age range in the normative sample and then estimate, or extrapolate, the AE scores for examinees at specified month intervals. This method of deriving AE scores is problematic because as age increases, there is a decrease in the differences in raw scores between age groups, making extrapolation for older examinees more difficult. Therefore, AE scores for higher raw scores are extrapolated for ages in much larger increments than for younger children who earn lower raw scores. This is evidenced in this study by the fact that the only two extrapolated AE scores within the 19-year-old age range were 19;1 and 19;9.

**LIMITATIONS OF THE STUDY AND FUTURE DIRECTION**

Although the results of this study provide evidence of the theoretical limitations associated with AE scores, some of the limitations mentioned earlier in this article could not be supported in this study. First, this study could not provide evidence to support the proposed limitation that AE scores do not provide specific information about an examinee’s performance on a given test. As was mentioned earlier, the examinee that earned an AE score of 14;0 may or may not have performed similarly to a 14-year-old. This limitation could be examined by administering a standardized test to subjects of varying ages and levels of cognition and by examining the performance patterns of examinees earning the same AE scores. Additional research is needed in this area to provide evidence of this limitation.

This study also did not provide evidence for the limitation that AE scores falsely imply that abilities increase at a constant rate from year to year. In contrast to standard scores, which follow an equal-interval scale, AE scores are ordinal, with a flattening of the curve as age increases (Aiken, 2003; Anastasi & Urbina, 1997; Lawrence, 1992; Salvia et al., 2006). This limitation may be investigated through a longitudinal study that examines the performance of a group of participants over time.

There are some additional shortcomings associated with this study that are worthy of note. First, this study included a relatively small group of participants. Second, the majority of the participants were college students and, therefore, may not have been reflective of the general population. Third, this study used the PPVT–III to examine the limitations of AE scores. This test is a measure of a very discrete skill—receptive vocabulary at the single word level. To generalize the findings of this study, additional research is needed to address these limitations. Specifically, future research should include larger sample sizes with participants of varying backgrounds (i.e., age, cognitive level, educational experiences). In addition, research should examine the use of AE scores to report performance on standardized tests that assess more diverse skills and behaviors.

In sum, the results of this study provide evidence for some of the limitations associated with AE scores that have been proposed in the literature. AE scores are highly ineffective in reporting performance on standardized tests. Yet, these scores continue to be used when explaining standardized test performance to parents and teachers and by agencies at the local, state, and federal level for qualifying children for special services. Although many of the limitations associated with AE scores have been described thoroughly in the literature, there was insufficient research that provided concrete evidence for these limitations. This study provides a starting point for research on the limitations of AE scores. Ideally, this research will lead to the replacement of AE scores with more useful and appropriate scores that take into consideration the range of normal performance, thus assisting SLPs in implementing evidence-based practice into the assessment process.

**REFERENCES**


Contact author: Emily S. Maloney, Center School, 49 Main Street, Ellington, CT 06029. E-mail: emsros@aol.com.