Special Education Distributions and Analysis

Valerie Felder
Department of Education, State of Michigan, felderv@michigan.gov

Shlomo S. Sawilowsky
Wayne State University, shlomo@wayne.edu

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation
Available at: http://digitalcommons.wayne.edu/jmasm/vol14/iss1/20

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.
Special Education Distributions and Analysis

Valerie Felder
Department of Education
State of Michigan

Shlomo S. Sawilowsky
Wayne State University
Detroit, MI

Micceri (1989) examined the distributional characteristics of 440 large sample general education achievement and psychometric measures. All the distributions were found to be statistically significantly different from the normal distribution. In this study, 395 special education datasets were examined. Although there were some normally distributed datasets, most were not, and some were markedly different in shape from those found by Micceri (1989). Implications for statistical testing and making special education policy decisions were given.

Keywords: Nonnormal data sets, statistical testing, special education

Special education distributions

Micceri (1989) conducted an investigation of the distributional characteristics of 440 large sample educational achievement and psychometric measures. The data sets were obtained from general education and the behavioral and social sciences, including ability tests, achievement tests, criterion or mastery level tests, psychometric measures, and pre- and post-intervention scores. All were found to be non-normal based on the Kolmogorov-Smirnov test with nominal $\alpha = 0.01$. Factors that contributed to a non-Gaussian error distribution in the population include (a) subpopulations within a target population, (b) ceiling/floor effects, and (c) variability in the items within a measure. This has implications in terms of statistical testing, because classical parametric tests require normality in order to maintain acceptable robustness and comparative power properties (Sawilowsky & Blair, 1992). If ignored, costly errors may occur in making policy decisions.

The prevalence of non-normally distributed data permeates many fields. Previous studies that demonstrated this include Bradley (1977, 1982), Hill and

Dr. Valerie Felder is Research Consultant with the Michigan Department of Education. Email her at: felderv@michigan.gov. Dr. Sawilowsky is a Professor in the Department of Evaluation and Research, College of Education. Email him at shlomo@wayne.edu.
Dixon (1982), Ito (1980), Pearson and Please (1975) and Tan (1982). However, they, as well as Micceri (1989), did not have special education and disability assessments as a focus.

Assessment of students in special education is frequently different than for students in general education, because often the focus is on process or progress as opposed to specific learning outcomes. This may include adaptive behavior, development, and screening. Adaptive behavior skills are those skills that are useful in daily functioning. Developmental skills pertain to fine- and gross-motor, communication and language, social, cognitive, and self-help skills. Screening helps find children who might be below the norm in different areas (Rosenberg, Westling, & McLeskey, 2010).

**Purpose of the study**

Given the paucity of representation of special education data sets in the studies mentioned above, the purpose of this study is to canvass that literature to determine the distributional shape commonly encountered. This will help inform the appropriate statistical method (i.e., parametric or nonparametric) to be used in measuring the progress of students in special education.

**Methodology**

The distribution patterns of special education data sets were obtained from published, peer-reviewed journal articles from the years of 2007-2011. In addition, research studies that focused on special education assessment were considered for inclusion. A Google Scholar search with the key terms “special education” and “data” returned 396,397 related publications.

To construct a confidence level of 95% and margin of error of ±5%, a sample size of 384 data sets was needed from that population. It was estimated a return response rate of 25% was needed to accommodate lack of responses, and therefore 1,540 survey requests were made from selected authors of those published studies. Assessment data sets were also solicited from various state departments of education. Requests were made via email and telephone. The request included instructions to de-identify student information. Initial contact via email and phone was made from October - December, 2012. Follow-up phone calls and email messages were made in January, 2013.
Criteria for inclusion
Potential studies were reviewed to determine if the instrument used to collect data was supported by adequate reliability and validity information. However, there was no preset type or minimum reliability index or validity methodology required for inclusion.

Reliability is “the consistency that a test measures whatever it measures” (Sawilowsky, 2007, p. 516). As noted by Sawilowsky (2000), reliability is a psychometric property of a test. If the test produces similar results under consistent conditions then it is considered reliable. There were different types of reliability information obtained:

- Internal consistency, which is the extent items on an instrument relate to each other.
- Test-retest, which is the consistency over time (i.e., stability) of an instrument.
- Inter-rater reliability, which is the degree of agreement among raters.

Validity is “the degree that a test measures what it purports to measure” (Sawilowsky, 2007, p. 166). There are different types of validity, including content-related validity, construct validity, and predictive validity (Cicchetti, 1994):

- Content-related validity, which is how well the content of the test relates to what is being assessed.
- Construct validity. “A construct is a fiction that is used to explain reality” (Cuzzocrea & Sawilowsky, 2009, p. 215), such as aptitude, intelligence, or self-determination. Hence, construct validity is the degree that a test measures that fiction used to explain reality.
- Predictive validity, which is the extent a test predicts some criterion measure.

Results
There were 744 authors contacted via email. Note that many authors had obtained multiple data sets in their study, exceeding the 1,540 data set requirement. Follow-up phone calls and emails were conducted where necessary after 3 months. There were n = 333 data sets collected from journal article authors, as compiled in
In addition, academic achievement special education assessment test scores were requested from state education departments. Twenty-four state departments of education, randomly selected, were contacted from which an additional \( n = 62 \) data sets were obtained from Alaska, Florida, Michigan, Minnesota, Missouri, and South Carolina, as compiled in Table 2. Thus, there were a total \( N = 395 \) data sets. Based on an estimated accessible population, the obtained sample size yielded a confidence level of 95% with a \( \pm 4.25\% \) margin of error.

**Table 1.** Summary of Canvassed Authors (744) and Data Sets (4,362)

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Total % of Articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptable Reliability</td>
<td>1760</td>
<td>40.30%</td>
</tr>
<tr>
<td>Acceptable Validity</td>
<td>1600</td>
<td>36.70%</td>
</tr>
<tr>
<td>Acceptable Articles*</td>
<td>1002</td>
<td>23.00%</td>
</tr>
<tr>
<td>Acceptable Data Sets</td>
<td>333</td>
<td>7.60%</td>
</tr>
</tbody>
</table>

*Note: An acceptable article required acceptable reliability and validity evidence.*

**Table 2.** Data Sets from State Departments of Education

<table>
<thead>
<tr>
<th>Department</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Florida</td>
<td>16</td>
</tr>
<tr>
<td>South Carolina</td>
<td>8</td>
</tr>
<tr>
<td>Missouri</td>
<td>3</td>
</tr>
<tr>
<td>Minnesota</td>
<td>19</td>
</tr>
<tr>
<td>Alaska</td>
<td>15</td>
</tr>
<tr>
<td>Michigan</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>62</strong></td>
</tr>
</tbody>
</table>

Cronbach alpha coefficients for the instruments used to obtain these data sets ranged from .70 to .93. Test-retest reliability coefficients ranged from .65 to .97, and alternate-forms reliability ranged from .91 to .92. Concurrent validity indices ranged from .70 to .89, and predictive validity indices ranged from .65 to .86. (The author of one study used Item response theory (IRT) in a measurement model (i.e., Rasch one-parameter logistic (1PL) partial credit model for polytomous scoring).

**Distribution shapes**

The histograms was analyzed and categorized. Histograms that resembled Micceri’s (1986) distributions were named accordingly. Histograms that did not
resemble Micceri’s distributions were given a name based on the shape of each distribution. Figure 1 contains typical shapes obtained from the data sets. The types of distributions and the percentage of each distribution that were collected are indicated in Table 3. There were 258 (65.31%) special education data sets that were different and 137 (34.67%) similar to Micceri’s (1989) shapes.

The data sets were also analyzed for normality and compared with Micceri’s data sets. Based on the Kolmogorov-Smirnov and Shapiro-Wilks tests, 318 (81%) data sets were non-normally distributed and 77 (19%) data sets were normally distributed. Recall that Micceri (1986, 1989) found 100% of the distributions to be significantly non-normally distributed at the $\alpha = 0.01$ level. There were 19 out of 440 distributions, or 4.3%, that were considered reasonable approximations to the Gaussian distribution only in the sense that they were smooth symmetric with light tails. As compared with Micceri’s (1986, 1989) results, this study shows special education assessment data sets were somewhat more likely to be normally distributed, but the number of different data sets shapes was higher than those found by Micceri (1986, 1989).

Table 3. Type, Number, and Percentage and Distribution Shapes

<table>
<thead>
<tr>
<th>Type of Distribution</th>
<th>Number</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extreme Bimodality</td>
<td>106</td>
<td>26.84%</td>
</tr>
<tr>
<td>Equimodal</td>
<td>96</td>
<td>24.30%</td>
</tr>
<tr>
<td>Unimodal and Smooth</td>
<td>79</td>
<td>20.00%</td>
</tr>
<tr>
<td>Bimodal and Smooth</td>
<td>31</td>
<td>7.85%</td>
</tr>
<tr>
<td>Slight Asymmetry</td>
<td>25</td>
<td>6.33%</td>
</tr>
<tr>
<td>Multimodal and Lumpy</td>
<td>19</td>
<td>4.81%</td>
</tr>
<tr>
<td>Unimodal and Slightly Smooth</td>
<td>10</td>
<td>2.53%</td>
</tr>
<tr>
<td>Extreme Asymmetry</td>
<td>6</td>
<td>1.52%</td>
</tr>
<tr>
<td>Slightly Asymmetric and Digit Preference</td>
<td>6</td>
<td>1.52%</td>
</tr>
<tr>
<td>Digit Preference</td>
<td>4</td>
<td>1.01%</td>
</tr>
<tr>
<td>Unimodal and Slightly Lumpy</td>
<td>4</td>
<td>1.01%</td>
</tr>
<tr>
<td>Equimodal and Symmetric</td>
<td>3</td>
<td>0.76%</td>
</tr>
<tr>
<td>Extreme Mass at Zero</td>
<td>2</td>
<td>0.51%</td>
</tr>
<tr>
<td>Mass at Zero</td>
<td>1</td>
<td>0.25%</td>
</tr>
<tr>
<td>Smooth Symmetric</td>
<td>1</td>
<td>0.25%</td>
</tr>
<tr>
<td>Equimodal and Slight Asymmetry</td>
<td>1</td>
<td>0.25%</td>
</tr>
<tr>
<td>Slightly Smooth and Symmetric</td>
<td>1</td>
<td>0.25%</td>
</tr>
</tbody>
</table>
Dataset 1. Skew = 2.090, PATM Pre-test

Dataset 2. Skew = 1.340, PATM Post-test

Dataset 3. Skew = -0.111, CAAVES Reading Assessment

Dataset 4. Skew = -0.080, CAAVES Math Assessment

Dataset 5. Skew = -0.246, Pre-test Tomlinson’s differentiated instruction strategies adapted assessment

Dataset 6. Skew = -1.543, Post-test Tomlinson’s differentiated instruction strategies adapted assessment

Dataset 7. Skew = 1.291
Grade 2, Dyslexia criteria, Spring

Dataset 8. Skew = 0.896
Grade 1, Fluency Word Recognition, Fall

Figure 1. Special Education Data Sets
Discussion

There were more classifications of special education data sets as extreme bimodality ($n = 106$, uni-modal, and smooth and equimodal than found in other disciplines. There were 106 extreme bimodality distributions and 57%, or 60 data sets, were non-normal. There were 46 distributions that were normal. There were 79 unimodal and smooth distributions and 29%, or 23 data sets, were non-normal. The remaining category, which had a large amount of distributions, is the equimodal category. There were 96 distributions and 70%, or 67, were non-normal. Thirty percent of the equimodal distributions were normally distributed based on the Kolmogorov-Smirnov and/or Shapiro-Wilks normality tests.

These data sets that were non-normally shaped pertained to curriculum-based assessments in writing, alternative assessments, applied problem solving, calculations, mathematics operations, reading, letter-word identification, segmenting words, and letter naming. Assessments of achievement, and fine- and gross-motor skills tended to be shaped normally.

In terms of policy, it is important to consider statistical robustness and comparative power when analyzing special education assessments. The results of this survey confirm the importance of considering nonparametric alternatives to parametric methods. As has been conducted throughout the Monte Carlo literature of the past century for data in many disciplines (e.g., general education, psychology, medicine, nursing), a study is warrant to determine the extent to which robustness and power of parametric tests may be compromised when analyzing special education data.

The new special education data shapes in this study may overlap with Micceri’s (1989) data shapes. Due to the small sample size of the special education data sets, some of the shapes were different than Micceri’s data shapes, but a larger sample sizes may show the data converges to one of Micceri’s shapes.

For example, consider the data sets from the Florida Alternate Assessment. They were separated by grade level and a distribution was created for each data set, because the achievement of students in special education is measured based on a set of academic standards for each grade level. However, if the sample size is increased by combining a single grade with all grade levels, the resulting shape, identified by Micceri (1989) as a discrete mass at zero with gape, will result, as noted in Figure 2.
In summary, Micceri’s (1989) seminal article on 440 real data sets from general education achievement and psychometric constructs, shockingly, found them all to be non-normally distributed. This led to a major overhaul in techniques for analyzing quantitative data, as is known in the statistical literature, in those fields. Unfortunately, progress in revising and updating statistical strategies into other fields has been slow. Workers have the tendency to hold fast to techniques learned many years prior in graduate school, and furthermore, with the uptick in qualitative research, the lessons learned from Micceri (1989) obtain little voice until such surveys are replicated in their fields. On the basis of 395 special education data sets obtained in this study, differences from Micceri’s (1989) rubric were noted, particularly the emergence of new non-normal distribution shapes. We believe this survey will help motivate quantitative workers in the special education field update their data analytic choices.
References


Developmental Disorders, 34(1), 7-17.
doi:10.1023/B:JADD.000018069.69562.b8


doi:10.1037/0033-2909.105.1.156


doi:10.1177/00131644021970439


Appendix: Journals used in the survey

Journals marked with an “*” were used in the survey. The data is available from the first author of this study.

*American Annals of Deaf
*American Educational Research Journal
*American Journal on Intellectual and Developmental Disabilities
*Annals of Dyslexia
*Applied Measurement in Education
Australasian Journal of Special Education
Behavioral Disorders
British Journal of Special Education
Career Development for Exceptional Individuals
Child Development Perspectives
Developmental Psychology
Early Childhood Research Quarterly
Education and Training in Mental Retardation and Developmental Disabilities
*Education and Treatment of Children
Educational Assessment
*Educational and Psychological Measurement
*Elementary School Journal
*Exceptional Children
*Exceptionality: A Research Journal
International Journal of Disability
*Journal of Adolescent and Adult Literacy
*Journal of Applied Behavior Analysis
Journal of Applied Developmental Psychology
Journal of the Association for Persons with Severe Handicaps
Journal of Attention Disorders
*Journal of Autism and Developmental Disorders
Journal of Deaf Studies and Deaf Education
*Journal of Disability Policy Studies
*Journal of Early Intervention
Journal of Educational Psychology
Journal of Educational and Behavioral Statistics
Journal of Educational Measurement
*Journal of Emotional and Behavioral Disorders
SPECIAL EDUCATION DISTRIBUTIONS AND ANALYSIS

*Journal of Intellectual Disability Research
*Journal of the International Association of Special Education
*Journal of Learning Disabilities
Journal of Policy and Practice in Intellectual Disabilities
*Journal of Positive Behavior Interventions
*Journal of Psychoeducational Assessment
Journal of Research and Development in Education
*Journal of School Psychology
*Journal of Special Education
Journal of Speech and Hearing Research
*Journal of Visual Impairment and Blindness
*Learning and Individual Differences
*Learning Disability Quarterly
*Learning Disabilities Research and Practice
Mental Retardation
Peabody Journal of Education
*Preventing School Failure
*Psychology in the Schools
*Reading and Writing
Reading Psychology
Reading Research Quarterly
*Remedial and Special Education
Research in Developmental Disabilities
*Review of Educational Research
*School Psychology Quarterly
*School Psychology Review
Teachers College Record
Teaching Exceptional Children
*Volta Review