Riverside Insights®

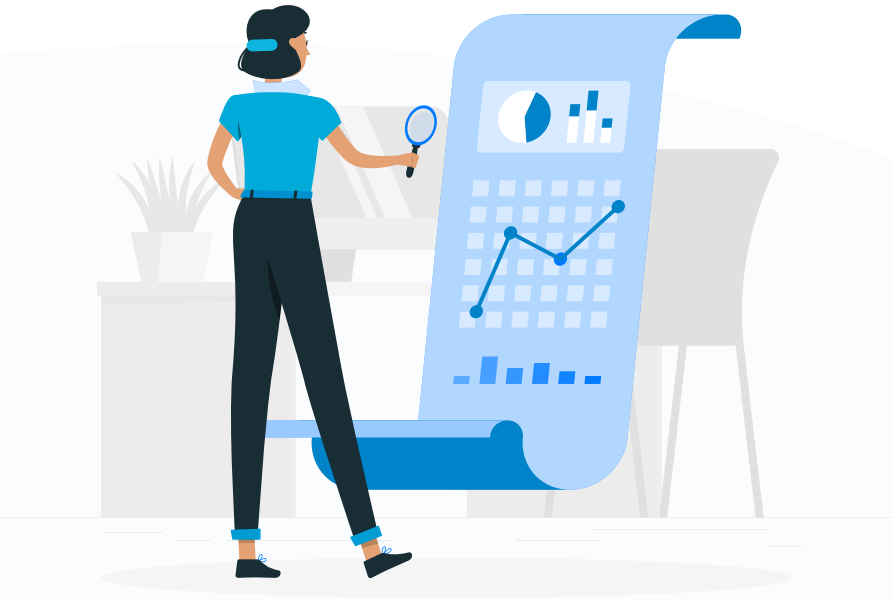
OLD VS NEW...

It is No "Apples to Apples" in the Assessment World

A Review of Differences Between Editions of the *Battelle® Developmental Inventory*

By Katy Genseke, Psy.D

In the world of assessment, there are tried-and-true measures which have always been used to gather information to guide decision making. The *Battelle Developmental Inventory* (BDI; Newborg et al., 1984) has been around since 1973 and is a staple in the assessment of developmental milestones for children birth 7 years 11 months. In September of 2020, Riverside Insights® released the 3rd edition of this important assessment. But why? And what are the differences between the two versions? Let us dig into that a bit more.



Why?

Although the industry typically revises and re-standardizes assessments every 10-12 years, the Standards for Educational and Psychological Testing clearly state that "Test specifications should be amended or revised when new research data, significant changes in the domain represented, or newly recommended conditions of test use may reduce the validity of test score interpretations" (American Educational Research Association [AERA], the American Psychological Association [APA], and the National Council on Measurement in Education [NCME], 2014). **This implies that assessments should be revised and re-standardized whenever significant changes in the research are made that make the test less reliable or valid, regardless of the passage of time.**

Test users are responsible for providing evidence that the older version is as appropriate as the new version for that particular test use.

When analyzing the BDI, one must take into consideration that in the years since the publication of the BDI-2® and the subsequent BDI-2® NU, there have been significant changes in the demographic characteristics of the U.S. population. Changes in population demographics, updated state standards, new research in early childhood development, and other factors such as a need for new, more child-friendly artwork have resulted in the third edition of the BDI. **The BDI-3™ is a substantive revision of the Normative Update and includes revised and expanded content, a new layout and art design, and a true contemporary normative sample that is representative of the predicted 2020 United States Census** (Newborg, J., 2020).

The Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014) further states, "If an older version of a test is used when a newer version has been published or made available, **test users are responsible for providing evidence that the older version is as appropriate as the new version for that particular test use.**" This is an important call out, as many practitioners continue to use older outdated assessments far longer than desirable, and find discrepancies among the data, therefore over-qualifying or under-qualifying children in certain categories, which is a disservice to the children being assessed.

Specific Differences Between the BDI-2 NU and the BDI-3

Demographic Differences

The enhanced normative data is designed to serve as a basis on which eligibility and placement decisions can reliably be made.

Test Blueprint

Test items were repositioned in different domains or subdomains, or in a different order within a subdomain, with greater attention to standards and guidelines put forth by states, federal programs, professional and other groups, and scientific researchers.

Test Content

With new item content and development, each subdomain now contains different content from the previous version that better align to new theories and research on the developmental abilities of children at each age level.

Item Scoring

Revisions were made to the item scoring criteria in the BDI-3 to add greater objectivity and precision to the scoring rubrics.

Score Differences

The tests are not directly comparable, and the new BDI-3 is better aligned to more accurately assess developmental milestones based on the most recent literature and data available.

Battelle Developmental Inventory 2nd edition NU Normative Update

Battelle3 Developmental Inventory 3rd Edition™

Specific Differences Between the BDI-2 NU and the BDI-3

Demographic Differences

The BDI-3 and BDI-2 NU standardization samples are composed of different examinees, tested approximately 15 years apart with relatively small differences in demographic composition. Most notable in changes is the continued shift, in terms of the percentage of the population, away from White demographic group toward other racial/ethnic groups. The decline in the percentage of examinees in the White group from BDI-2 NU to the BDI-3 is evident with the increase in the combined percentages of the Asian and Other/Mixed race groups. **The enhanced normative data is designed to serve as a basis on which eligibility and placement decisions can reliably be made**. The updated standardization data also form the basis for establishing the reliability of the scores as measures of student level and progress.

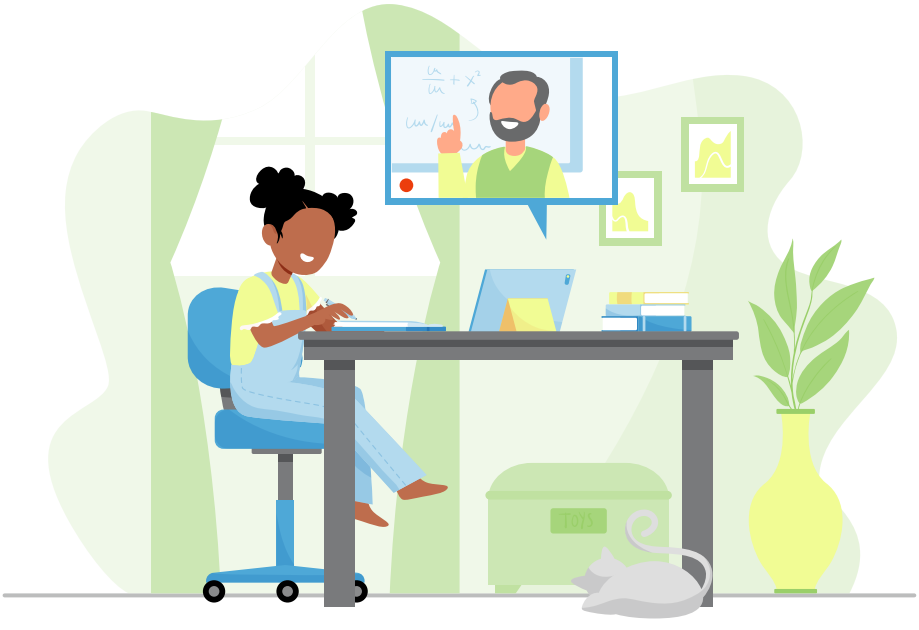
Test Blueprint

The number of subdomain scores which contribute to the overall domain score has changed from the BDI-2 NU to the BDI-3. Overall, **test items were repositioned in different domains or subdomains, or in a different order within a subdomain, with greater attention to standards and guidelines put forth by states, federal programs, professional and other groups, and scientific researchers**.

Specific Differences Between the BDI-2 NU and the BDI-3

Test Content

As mentioned above, there are many changes to the test content based on new research over the past 15 years. Revised items present information more consistently across administration procedures and better align with scoring guidelines. Additional items and increased upper age ranges for several subdomains (Self-Care, Adult Interaction, Peer Interaction, Gross Motor, Fine Motor, and Attention and Memory) have also been included. **With new item content and development, each subdomain now contains different content from the previous version that better align to new theories and research on the developmental abilities of children at each age level.** For example, for children ages 6 years and older, they now can complete all 13 subdomains, while in the BDI-2 NU, they could only complete 7. In the Adaptive Domain, the BDI-3 Self-Care Subdomain includes a greater balance of items oriented toward eating, toileting, dressing, and sleeping. The BDI-3 Social-Emotional Domain (previously the Personal-Social Domain) was revised and reorganized to reflect the current social-emotional CASEL standards. In the Communication Domain, the Expressive Communication Subdomain was expanded to address speech articulation more completely.



Specific Differences Between the BDI-2 NU and the BDI-3

Item Scoring

The three-point objective scoring system provides for a sensitive measure that considers the continuum of development for a given milestone—from not observed, to emerging, to fully developed. **Revisions were made to the item scoring criteria in the BDI-3 to add greater objectivity and precision to the scoring rubrics**. The scoring rubrics are written in a concrete and concise manner, making it easy for the examiner to choose from the score categories. The scoring rubrics are also better aligned to the administration procedures so that users have the information needed to score the items.

The administration procedures are designed to collect data through a structured test format to assess the child; observations of the child in natural settings; and interviews with parents, caregivers, teachers, and/or other adults familiar with the child. **The availability of multiple data sources allows a more complete evaluation of a child's functional abilities that takes into account what the child can do on command with the examiner; behaviors observed in his or her natural environment; or reports by a parent, caregiver, or teacher and is consistent with legislative mandates for multidimensional assessment**.

The BDI-3 also offers online scoring and reporting through the *Riverside Score* platform, as well as digital administration through the Mobile Data Solution (MDS) for free with the BDI-3 subscription. These features have been noted as absolutely essential in ensuring the fidelity of data. Many times, practitioners who hand score make small mistakes in tallying raw scores or converting raw scores to standardized scores to report out performance. These mistakes lead to misdiagnoses more often than most realize. **Even a one-point scoring error can be the difference between someone receiving services or being denied services**. This line is especially sharp when using any form of ability or achievement assessment for students being considered for specific learning disabilities (SLD), based on cut-off scores. The legitimacy of these high-stakes decisions depends entirely on the accuracy of test scores (McDermott et al., 2014). **By using the guided administration and automatic scoring features of the MDS and Riverside Score platform, fidelity of data and decision making is more accurate and precise**. States who are responsible for reporting out their data to the federal government can trust the data from the Riverside Score platform much more than data hand scored and entered into a database.

Specific Differences Between the BDI-2 NU and the BDI-3

Score Differences

Score differences are most notable by practitioners who have used the previous version of any given assessment. If an examiner were to administer the BDI-2 NU and the BDI-3 to the same examinee, score differences would be seen between the two administrations.

When any new version of an assessment is released, there will be score differences based on the changes in content, items, art, and metrics, as well as changes within the standardization population. **It is important to realize you are not comparing "apples to apples" when revisions are published**.

Research has been conducted on the differences between the BDI-2 NU and the BDI-3 scores to help practitioners understand what to be aware of when transitioning to the new version. A full account of this data is published in the Examiner's Manual, as well as in this Assessment Service Bulletin.

In general, more significant score differences may be noted within the BDI-3 in domains such as the Cognitive Domain, as measurements which assess cognitive ability typically see more significant changes. It has been observed that the average score of populations on cognitive tests increases over time (Graham & Plucker, 2001-02). The researcher who discovered this phenomenon is James R. Flynn and it is referred to as the "Flynn Effect." The Flynn Effect has been observed consistently in every country it has been studied in across the world (Flynn, 1987). It is because of the Flynn Effect, this increase in population intelligence test scores over time, that cognitive tests must be periodically re-standardized.

Specific Differences Between the BDI-2 NU and the BDI-3

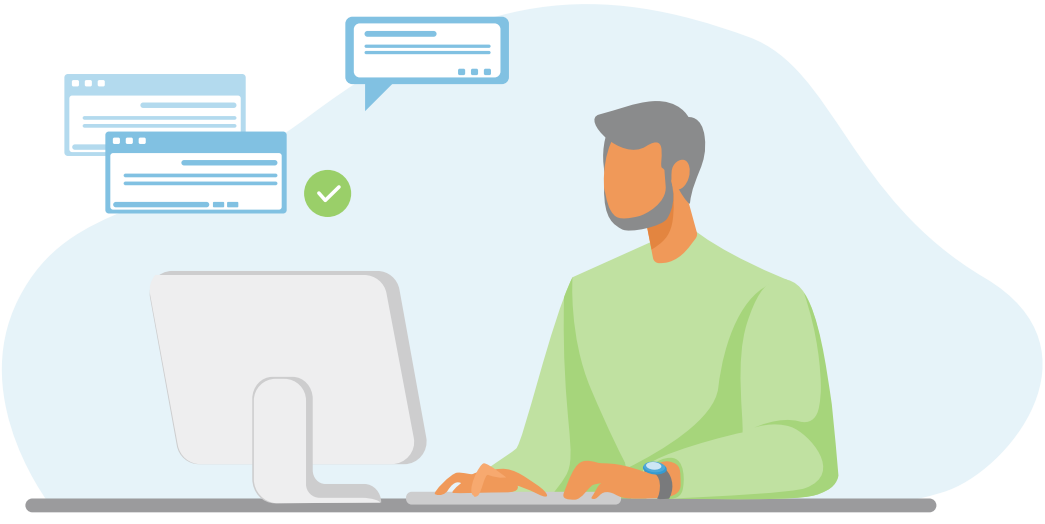
Examiners will typically see score differences in all domains. The magnitude of score differences will be seen at the domain level, as opposed to the subdomain level. **Overall, score differences suggest children are expected to obtain domain standard scores that are slightly higher with the BDI-3 than with the BDI-2 NU**. However, at the domain level 94% of the scores are within one third of a standard deviation of their respective score metric and are considered small or less than small psychometrically.

This phenomenon is very important to understand, as many examiners become confused when they see score differences and believe they should be able to compare "apples to apples." However, **with changes in content, scoring, item alignment, standardization populations, and blueprint it becomes understandable that changes in scores are inevitable**. For this reason, **examiners are cautioned against using the BDI-2 NU to assess entrance eligibility, as the new version, which is better aligned to our current population and research is the best tool to determine early intervention and/or special education eligibility**. Additionally, assessing the whole child is critical in identifying strengths, weaknesses, and possible eligibility into early intervention or special education programs. **It is essential to gather multiple points of data before ever diagnosing a child with a disability**.

The Ethical Role of a Practitioner

It is important that the BDI-3 be administered and interpreted by people who have appropriate training and competence to prevent misuse of the test. If the potential BDI-3 user recognizes his or her limitations in specific skills or knowledge areas that appear to be relevant to the assessment, he or she should find a colleague or other resource (such as trainings) that can assist in developing the necessary skills or an understanding of the relevant technical information. **BDI-3 users must be grounded in the instrument's critical features, particularly the standardization sample, administration procedures and scoring, and calculation and explanation of various scores**. Training on demand is available through the Riverside Training Academy.

Overall, it is important to know why new editions of tests are constructed and why it is important to use the most up to date assessments when making high stakes eligibility determinations. Before using any new version, make sure you read up on the changes made to the assessment, the reasons the changes were made, how those changes impact the interpretation of the data, and of course get trained on the proper administration of the new version. **There are no "apples to apples" with new tests, so be careful when comparing student progress based on two different versions**. And remember, the best data is gathered by those who are well trained not just in the administration but in the interpretation.



References

- American Educational Research Association (AERA), American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. AERA.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101,171-191.
- Graham, C. & Plucker, J (2001-02). The Flynn effect. Retrieved March 15,2021 from http://www.indiana.edu/~intell/flynneffect.shtm
- McDermott, P. A.,Watkins, M.W., & Rhoad, A.M. (2014).Whose IQ is it? Assessor bias variance in high-stakes psychological assessment. Psychological Assessment, 26(1), 207–214.
- Newborg, J., Stock, J., Wnek, L., Guidubaldi, J., & Svinicki, J. (1984). Battelle Developmental Inventory: Examiner's manual. DLM/Teaching Resources.
- Newborg, J. (2020). Battelle Developmental Inventory, 3rd Edition: Examiner's manual. Riverside Assessments, LLC.