NAEP DBA-PBA Linking Error With EdSurvey

Paul Bailey[*][†]

July 1, 2021

Introduction

Students who take paper-based assessments (PBAs) and digitally based assessments (DBAs) may experience items differently—for example, an item may be more or less difficult in the DBA format than in the PBA format. To account for this, the National Center for Education Statistics (NCES) drew two nationally representative samples and administered one as a PBA and the other as a DBA. Because the samples were both nationally representative, the overall mean can be assumed to be equivalent, and the two samples were linked under stochastic equivalence. However, this linking introduces additional variance and the *linking error* accounts for the additional variance, from the change in modes. The methodology assumes that the mode effect is the same for the entire sample and accounts for the variance introduced resulting from to the change in mode.

`EdSurvey` implements the method, developed by NCES and described in this document, to account for linking error in assessments that have combined PBA and DBA formats. As of June 2021, this includes 2018 Grade 8 Civics, Geography, and U.S. History; 2019 Science; and 2019 Grade 12 Mathematics and Reading. This method is not relevant to 2017 Grade 4 and Grade 8 Mathematics and Reading. `EdSurvey` does implement linking error for these studies, but it is not described in this document.

Three groups of plausible values (PVs) are used to estimate statistics and capture linking error in variance estimation:

- A first group of PVs is used to estimate the quantity of interest.
- A second group of PVs is used to estimate sampling variance.
- A third group of PVs is used to estimate imputation variance.

NCES distributes data from the National Assessment of Educational Progress (NAEP) with only the first group of PVs present, so the second and third groups must be calculated before the linking error variance can be estimated. All three groups exist for each subject scale (e.g., composite, univariate) and each subscale (when present). Although the theta values are not used directly by the user, they are used in intermediate steps and appear on the files. The theta scores available for use directly in `EdSurvey`.

This document first describes using `EdSurvey` to calculate the linking error and then the formulas `EdSurvey` uses to calculate linking error.

In EdSurvey

Beginning with `EdSurvey` 2.7, the NAEP linking error formulas have been implemented in most of the `EdSurvey` analytical functions including `edsurveyTable`, `achievementLevels`, `percentile`, `gap`, `cor.sdf`, `rq.sdf`, `lm.sdf`, `logit.sdf`, `probit.sdf`, and `waldTest` (when applied to the results of these methods), but

not in `summary2`, `mixed.sdf`, `mml.sdf`, or `mvrlm.sdf`. Users of `EdSurvey` can account for the linking error simply by using `_linking` outcomes in place of the subject scale or subscale. They also are listed separately when `showPlausibleValues` is called. For example, a table can be generated with or without linking error as follows.

```
require(EdSurvey) # must be 2.7.0 or higher
math12 <- readNAEP("path/to/Data/M50NT3AT.dat")
showPlausibleValues(math12)
# Results, not accounting for DBA-PBA linking
edsurveyTable(composite ~ dsex, data=math12)
# Results, accountint for DBA-PBA linking
edsurveyTable(composite_linking ~ dsex, data=math12)
```

There are a few limitations of the linking error procedure. First, it does not allow Taylor series variance estimation. Second, the value `jrrIMax` must always be one. Attempting to use `EdSurvey` with Taylor series variance estimation or `jrrIMax` greater than one will result in an error.

Also, becuase all the necessary variables (shown in the next section) are not provided by NCES in the NAEP data, reading in data is slowed by computing the additional intermediate and linking error variables (often taking several minutes), and calculations are slower when the linking error is calculated.

Formulas

This section shows the formulas used by `EdSurvey` to compute statistics while accounting for linking error.

Estimation

A statistic is estimated using the usual scale score PVs by calculating the statistic with the full sample weights and averaging the statistic across the PVs.

$$Q = \frac{\sum_{j=1}^{J} Q_j}{J} \tag{1}$$

where Q is the estimated statistic; Q_j is the statistic estimated with the jth set of PVs, of which there are J; and all estimates are weighted results from the full sample weights (`origwt` on NAEP).

Overview of NAEP Linking Error Methodology

For this exercise, we will use the DBA student's theta scores and scale them to the PBA reporting scale with an equation of the form

$$X = b + a\theta \tag{2}$$

where X is the scale score, b is the location parameter, a is the scale parameter, and θ is the theta score. Although it is possible to map scale scores directly from one reporting scale to another, that more complicated procedure is neither described nor needed for this exercise.[1]

The values a and b are chosen so that students who took the DBA will have their θ scores mapped to the same mean and standard deviation as students who took the PBA. This scaling is performed by imputation index for the imputation variance and by replicate weight for the sampling variance, as described in the next two sections.

---

[1] Although 'EdSurvey' transforms DBA theta scale PVs to the the PBA reporting scale for linking error estimation, the NAEP data companion describes an equivalent method that maps the DBA reporting scale to the PBA reporting scale instead. As such, the transformation coefficient tables provided in the NAEP data product also reflect the transformation from the DBA reporting scale to the PBA reporting scale. The two methods are equivalent because both would yield identical transformed DBA PVs (on the PBA reporting scale) as intermediates for subsequent linking error estimation.

As is typical with the NAEP, scores are scaled by subscale and then the composite score is a weighted combination of

$$X_c = \sum_{k=1}^{K} \beta_k \cdot X_k \tag{3}$$

where X_c is the composite scale score, β_k is the weight applied the the kth subscale, and X_k is the kth subscale scale score. Note that the sum of the β values is always one. If there are no subscales, the overall (sometimes also called univariate) score can be used in the following sections and the number of subscales is simply one.

Students who took the DBA will have the variable `dbapba` set to `"DBA"` and students who took the PBA will have the variable set to `"PBA"`.

Imputation Variance   Ideally, the imputation variance would be summed sumed over all possible combinations of the 20 PBA and 20 DBA imputations, which is far more than can be reasonably calculated. To minimize computation time, only five permutations of the PVs are chosen; they are shown in Table 1, where the row number is the index of the DBA PV to use and the cell value is the index of the PBA PV to use.

The data already have the PVs used for estimation on it but need the imputation variance and sampling variance PVs added.

The algorithm for generating the imputation variance PVs is as follows:

- Define w_{DBA} as the vector of weights for students who took the DBA. The vector has length n_{DBA}.
- Define w_{PBA} as the vector of weights for students who took the PBA. The vector has length n_{PBA}.
- For each column (n) of the permutation table
  - For each row (j) of the permutation table
    * For each subscale (k)
      · Define X_k as the vector of the kth subscale, scale score PV for all students who took the PBA. Select the PV based on the element in Table 1 for this row and column (the j, nth element from the permutation table).
      · Define θ_k as the vector of the kth subscale, theta score j (the jth theta score PV) for all students who took the DBA.
      · Define $\mu_t = \dfrac{\sum w_{DBA}\theta_k}{\sum w_{DBA}}$
      · Define $s_t = \sqrt{\dfrac{\sum w_{DBA}(\theta_k - \mu_t)^2}{\sum w_{DBA}}}$
      · Define $\mu_x = \dfrac{\sum w_{PBA}X_k}{\sum w_{PBA}}$
      · Define $s_x = \sqrt{\dfrac{\sum w_{PBA}(X_k - \mu_x)^2}{\sum w_{PBA}}}$
      · Generate a new scale score for the DBA students, centered at the mean for the PBA students using $X'_k = b + a\theta_k$ where $a = \frac{s_x}{s_t}$ and $b = \mu_x - a\mu_t$.
      · The j, nth *imputation variance PV* for subscale k is X'_k for students who took the DBA and X_k for students who took the PBA.
    * Generate the composite for DBA students using the composite formula $X'_c = \sum_k \beta_k X'_k$.
    * The j, nth *imputation variance PV* for the composite is X'_c for students who took the DBA and X_c for students who took the PBA.

Table 1: Permutation Table

Row	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5
1	1	14	6	2	19
2	2	11	4	14	10
3	3	18	3	7	13
4	4	15	9	9	8

Row	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5
5	5	20	11	20	17
6	6	17	10	13	12
7	7	3	2	16	1
8	8	16	17	8	11
9	9	12	19	4	5
10	10	6	13	11	6
11	11	19	7	19	18
12	12	2	8	1	9
13	13	1	15	18	15
14	14	13	18	10	2
15	15	5	20	6	3
16	16	7	12	12	7
17	17	4	5	15	14
18	18	9	16	3	20
19	19	8	14	5	16
20	20	10	1	17	4

The above steps need to be performed only once. Then, for each statistic the imputation variance is calculated as follows. Using the imputation variance PVs, the linking imputation variance is then calculated per permutation (column) and then averaged acrosss the permutations.

- For each column (n) of the permutation table
    - For each row (j) of the permutation table, calculate the statistic in question using the j, nth imputation variance PV and the full sample weights, call this Q_{jn}.
    - Calculate $Q_{\cdot n} = \frac{1}{J} \sum_{j=1}^{J} Q_{jn}$ where J is the number of permutation table rows.
    - Calculate $V_{imp,n} = \frac{J+1}{J(J-1)} \sum_{j=1}^{J} \left(Q_{jn} - Q_{\cdot n} \right)^2$.
- Calculate $V_{imp} = \frac{1}{N} \sum_{n}^{N} V_{imp,n}$ where N is the number of columns in the permutable table

The value V_{imp} is the estimated imputation variance for the statistic in question.

Sampling Variance   Then generate the sampling variance PVs, using only the first PV but reweighting using the replicate weights.

- For each subscale (k)
    - Define X_k as the vector of the kth subscale, scale score PV for all students who took the PBA. Select the first PV.
    - Define θ_k as the vector of theta scores for the kth subscale for students who took the DBA. Select the first PV.
    - For each replicate weight (i)
        * Define $w_{DBA}^{(i)}$ as the vector of the ith replicate weights for students who took the DBA. The vector has length n_{DBA}.
        * Define $w_{PBA}^{(i)}$ as the vector of the ith replicate weights for students who took the PBA. The vector has length n_{PBA}.
        * Define $\mu_t^{(i)} = \frac{\sum w_{DBA}^{(i)} \theta_k}{\sum w_{DBA}^{(i)}}$
        * Define $s_t^{(i)} = \sqrt{\frac{\sum w_{DBA}^{(i)} (\theta_k - \mu_t^{(i)})^2}{\sum w_{DBA}^{(i)}}}$
        * Define $\mu_x^{(i)} = \frac{\sum w_{PBA}^{(i)} X_k}{\sum w_{PBA}^{(i)}}$
        * Define $s_x^{(i)} = \sqrt{\frac{\sum w_{PBA}^{(i)} (X_k - \mu_x^{(i)})^2}{\sum w_{PBA}^{(i)}}}$

* Generate a new scale score for the DBA students, centered at the mean for the PBA students using $X_k'^{(i)} = b^{(i)} + a^{(i)}\theta_k$ where $a^{(i)} = \frac{s_x^{(i)}}{s_t^{(i)}}$ and $b^{(i)} = \mu_x^{(i)} - a^{(i)}\mu_t^{(i)}$.

* The ith *sampling variance PV* is then $X_k'^{(i)}$ for students who took the DBA and X_k for students who took the PBA.

As previously, these steps need to be performed only once. Then for each statistic, the sampling variance is calculated as follows. Using the sampling variance PVs, the linking sampling variance is then calculated per replicate weight and summed.

- For each replicate weight (i), calculate Q_i as the statistic in question using the ith sampling variance PV and the ith replicate weight.
- Calculate the mean $Q. = \frac{1}{I}\sum_{i=1}^{I} Q_i$, where I is the number of replicate weights.
- Calculate $V_{samp} = \sum_{i=1}^{I} (Q_i - Q.)^2$.

The value V_{samp} is the estimated sampling variance for the statistic in question.