NEPS SURVEY PAPERS

Sabine Zinn, Ariane Würbach, Hans Walter Steinhauer,

and Hans Kiesl

THE COMPOSITE GRADE 4 WEIGHT OF THE KINDERGARTEN COHORT OF THE NATIONAL EDUCATIONAL PANEL STUDY

NEPS Survey Paper No. 44
Bamberg, August 2018

LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

NEPS
National Educational Panel Study

Survey Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS Survey Paper Series provides articles with a focus on methodological aspects and data handling issues related to the German National Educational Panel Study (NEPS).

The NEPS Survey Papers are edited by a review board consisting of the scientific management of LIfBi and NEPS.

They are of particular relevance for the analysis of NEPS data as they describe data editing and data collection procedures as well as instruments or tests used in the NEPS survey. Papers that appear in this series fall into the category of 'grey literature' and may also appear elsewhere.

The NEPS Survey Papers are available at https://www.neps-data.de (see section “Publications“).

Editor-in-Chief: Corinna Kleinert, LIfBi/University of Bamberg/IAB Nuremberg

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg − Germany − contact@lifbi.de

The Composite Grade 4 Weight of the Kindergarten Cohort of the National Educational Panel Study

Sabine Zinn[1], Ariane Würbach[1], Hans Walter Steinhauer[1], & Hans Kiesl[2]

[1]*Leibniz Institute for Educational Trajectories*
[2]*Ostbayerische Technische Hochschule Regensburg*

Technical Report referring to DOI:10.5157/NEPS:SC2:7.0.0

E-mail address of lead author:

methoden@lifbi.de

Bibliographic data:

Zinn, S., Würbach, A., Steinhauer, H. W., & Kiesl, H. (2018). *The Composite Grade 4 Weight of the Kindergarten Cohort of the National Educational Panel Study* (NEPS Survey Paper No. 44). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

The Composite Grade 4 Weight of the Kindergarten Cohort of the National Educational Panel Study

Abstract
This report details the derivation of the joint Grade 4 weight for the pooled samples of the NEPS Starting Cohort 2 (SC2). In total, the SC2 comprises three subsamples: the sample of Kindergarten children drawn in 2010, a refreshment sample of Grade 1 students drawn in 2013, and a sample of children being part of both. The two latter samples have been surveyed in the school context from the Grades 1 to 3. In contrast, only the parents of the first sample have been surveyed (via telephone) in this time. Since Grade 4 the children of all of the samples are surveyed again. Pooled analysis of the related data requires a joint weight for all of the children being part of the grade 4 sample. This reports details the corresponding weighting procedure. Concretely, we constructed a composite weight by minimizing the variance of a weighted sum of population sizes. All in all, the composite weight refers to the Grade 4 student population in the school year 2014/15.

1 Prequel

This report details the derivation of the (joint) Grade 4 / Wave 6 weight for the pooled samples of the Starting Cohort 2 (SC2). It supplements the previous and current nonresponse and weighting reports by Würbach (2018a, 2018b) which detail the nonresponse processes so far observed in the SC2 samples together with the nonresponse adjusted design weights derived.

The National Educational Panel Study (NEPS) surveys a cohort sample of Kindergarten children and Grade 1 students (Starting Cohort 2) and follows them over their transition to elementary school and beyond. The data are released via corresponding Scientific Use Files (SUF). The current SUF version is available under DOI:10.5157/NEPS:SC2:7.0.0.[1]

The initial SC2 sample was drawn in German Kindergartens in 2010. In 2013 (Wave 3), the sample of the Kindergarten children transitioned to elementary school. Children who transitioned to schools that were previously used to construct the Kindergarten sample (by an indirect sampling method, cp. Kiesl (2016) or Steinhauer, Aßmann, Zinn, Goßmann, and Rässler (2015)) were followed up in these schools. Subsequently, we denote this sample as *KIGA_PANEL*. Besides that, the SC2 sample was augmented by the classmates of the *KIGA_PANEL* children and another school sample yielding a further SC2 sample which we denote as *K*1*_AUF*, cp. Steinhauer, Zinn, Gaasch, and Goßmann (2016). From Grade 1 (in 2013, Wave 3) to Grade 4 (in 2015/16, Wave 6) the two samples *KIGA_PANEL* and *K*1*_AUF* were tested and surveyed together in their schools. All children of the initial SC2 sample who are not part of *KIGA_PANEL* constitute the third SC2 sample *KIGA_IND*. Over the period 2013-2015 (Waves 3-5), these children were tracked individually but neither surveyed nor tested. Only their parents were contacted and interviewed on the telephone.[2] Then, in Wave 6 (Grade 4) the entire sample was surveyed and tested again.

Detailed information on sample sizes, nonresponse process, and attrition patterns is given in Würbach (2018a, 2018b), Steinhauer et al. (2016) and Zinn, Würbach, Steinhauer, and Hammon (2018).

2 Composite Weighting

Multi-purpose analyses may require one weight for the pooled Grade 4 sample comprising the three SC2 subsamples in Wave 6. A straightforward way to derive such a weight is building a composite weight (Chu, Brick, & Kalton, 1999) from the Wave 6 weights of the SC2 subsamples.[3] Depending on the population parameter used for its construction, several such weights may arise. Mind that the population of the SC2 Wave 6 subsamples comprises Grade 4 students in the school year 2015/16. Without loss of generality, we use the size of this population as

[1] For general information on the NEPS, see Blossfeld, Roßbach, and von Maurice (2011). More detailed information is available in the documentation section on the homepage.

[2] Also the parents of the *KIGA_PANEL* and *K*1*_AUF* were interviewed on the telephone.

[3] An alternative to this method is to compute first weighted survey estimates for each sample. Then, these survey estimates are combined by a weighting average, where the weighting factors of the weighted average minimize the variance of the combined estimate. This method is theoretically sound but cumbersome to implement, especially when several survey estimates are considered (Chu et al., 1999).

the population parameter of interest to construct the joint Grade 4 weight.[4] Here, elementary schools are used as defined by Statistisches Bundesamt (2016).

Subsequently, we substantiate and detail the construction strategy of the joint weight for the pooled Grade 4 / Wave 6 SC2 data sample. Let n^A denote the sum of the sample sizes of *K1_AUF* and *KIGA_PANEL* directly before the Wave 6 survey (i.e., the number of children asked for participating in the Wave 6 tests and survey) and n^B the respective sample size of *KIGA_IND*. Denote further w^A as the joint nonresponse adjusted and calibrated survey weight before Wave 6 for *K1_AUF* and *KIGA_PANEL* and w^B as the related survey weight for *KIGA_IND* (i.e., the panel entry weights for Wave 6). Both sets of weights are part of the weights data provided within the SUF SC2:7.0.0 and described in very detail in Würbach (2018a). Table 1 in Section 3 gives their summary statistics. Having w^A and w^B at hand we can derive unbiased estimates of the true population size N:

$$\widehat{N}^A = \sum_{i=1}^{n^A} w_i^A = 705{,}396 \quad \text{and} \quad \widehat{N}^B = \sum_{i=1}^{n^B} w_i^B = 705{,}396$$

with

$$n^A = 6661 \quad \text{and} \quad n^B = 2383.$$

Thus, for any choice of α with $\alpha \in (0,1)$ also

$$\widehat{N}_\alpha = \alpha \sum_{i=1}^{n^A} w_i^A + (1-\alpha) \sum_{i=1}^{n^B} w_i^B$$

is an unbiased estimate of N. As a direct consequence, a joint weight $w = [w_1, \ldots, w_{n^A+n^B}]$ for the Grade 4 students (directly before the Wave 6 survey) computes as

$$w_i = \begin{cases} \alpha w_i^A & \text{for all students } i \text{ who are part of } K1\_AUF \text{ or } KIGA\_PANEL \text{ and} \\ (1-\alpha) w_i^B & \text{for all students } i \text{ who are part of } KIGA\_IND \end{cases}$$

for $i = 1, \ldots, n^A + n^B$. This weight can be applied to all kinds of weighted survey statistics (not only to linear ones) since $\sum_i w_i^A = \sum_i w_i^B = 705{,}396$. For more details see Chu et al. (1999).

To minimize the design effect of the weighted estimate (i.e., to avoid unnecessary variance inflation because of weighting), α has to be chosen to minimize the variance of the estimated population parameter, here, the population size $\widehat{N}_\alpha$. The variance of $\widehat{N}_\alpha$ is

$$Var(\widehat{N}_\alpha) = \alpha^2 Var(\widehat{N}^A) + (1-\alpha)^2 Var(\widehat{N}^B) + 2\alpha(1-\alpha) Cov(\widehat{N}^A, \widehat{N}^B).$$

The derivative with respect to α gives

$$\frac{\partial Var(\widehat{N}_\alpha)}{\partial \alpha} = 2\alpha Var(\widehat{N}^A) - 2(1-\alpha) Var(\widehat{N}^B) + 2(1-2\alpha) Cov(\widehat{N}^A, \widehat{N}^B).$$

Setting this equation to zero yields the following optimal value for α (i.e., the α value that gives

[4] Children leaving Kindergarten earlier or staying longer in Kindergarten (i.e., not entering school in the school year 2012/13) as well as students repeating and skipping a grade are not considered in the weighting procedure. In sum, they constitute 4.8 percent of the pooled sample.

the minimal variance):

$$\alpha = \frac{Var(\widehat{N}^B) - Cov(\widehat{N}^A, \widehat{N}^B)}{Var(\widehat{N}^A - \widehat{N}^B)} \,.$$

Beware that no value of α can be optimal for all kinds of analyses. Using the population size as the population parameter to derive α, targets all weighted analyses to the whole population of Grade 4 students. Considering subgroups of the Grade 4/ Wave 6 sample may yield distinct optimal values for α and thus smaller variances. In this case, the NEPS data users are asked to compute the optimal α for the specific subgroup considered by themselves. following the procedure described in this report.

Because the *K*1*_AUF* sample has been drawn nearly independently from the *KIGA_IND* sample[5], both samples can be considered as being approximately independent.

After 6 Waves, the probability of being part of the SC2 sample is mainly driven by the childrens' response behavior. We see that after such a long time the participation behavior of the *KIGA_PANEL* and *K*1*_AUF* children is very similar, whereas it significantly differs from the participation behavior of the *KIGA_IND* children, see Würbach (2018a). To a very high degree the childrens' participation propensities depend on whether they are tested/ surveyed at school (in groups) or at home (individually). Therefore, the dependency between the combined sample *K*1*_AUF* & *KIGA_PANEL* and the sample *KIGA_IND* can be considered as being (negligibly) small. Thus, also these both samples can be considered as being almost independent of each other.

Thus, $Cov(\widehat{N}^A, \widehat{N}^B) \approx 0$. The formula for the optimal α value simplifies to

$$\alpha \simeq \frac{Var(\widehat{N}^B)}{Var(\widehat{N}^A) + Var(\widehat{N}^B)} \,.$$

Note that this formula can also be applied to population parameters that are distinct from the population size.

When estimating the variances $Var(\widehat{N}^A)$ and $Var(\widehat{N}^B)$ we have to consider that the SC2 children have unequal (adjusted) design weights. Neglecting this feature may lead to biased variance estimates. An estimator that shows good properties in large samples, such as the SC2 ones, is (Wolter, 2007, p. 336):[6]

$$Var(\widehat{N}_\alpha) = \frac{1}{n_k(n_k-1)} \sum_{i=1}^{n_k} \left(w^k - \widehat{N}^k\right)^2, \quad \text{for } k \in \{A, B\} \,.$$

We yield $Var(\widehat{N}^A) = 74{,}689{,}810$ and $Var(\widehat{N}^B) = 208{,}717{,}929$. The optimal value for the corresponding α is 0.736. This value indicates a high preference for the weights of the *K*1*_AUG* & *KIGA_PANEL* sample compared to the *KIGA_IND* weights whose variance is notably higher.

[5] Both samples are only nearly independent because the 212 schools used in the indirect sampling procedure for the Kindergarten children in 2010 were also part of the school sampling frame for the *K*1*_AUF* sample which comprised 16,824 elementary schools, cp. Steinhauer et al. (2016) and Hellrung, Bockelmann, Schneider, Waschk, and Hillen (2013).

[6] Here applied to the population size as population parameter of interest.

3 Summary of Joint Grade 4 Weight

The NEPS provides various kinds of weights for Kindergarten children and elementary school students together with design information. These weights are described in detail in Würbach (2018a) and can be found in the SUF files `SC2_WeightsKindergarten_7-0-0`, or `SC2_WeightsElementarySchool_7-0-0`. This report deals with the joint weight for Grade 4 students before the Wave 6 survey (i.e., for all students who were asked to participate in Wave 6). In the SUF weights file `SC2_WeightsAsofGrade4_7-0-0`, this joint weight is denoted as `w_p6_joint`. Table 1 provides some summary statistics of `w_p6_joint`. Additionally, summary statistics of the related weights w^A and w^B of the subsamples *K1_AUF* & *KIGA_PANEL* and *KIGA_IND* are given (in the SUF weights file denoted as `w_p6`). The joint Grade 4 weight can be used in standard statistical analyses without further ado. General advices concerning the usage of weights are given in Würbach (2018a).

Table 1: Summary statistics for all weights provided.

Label of weight	**Min.**	**Lower Quart.**	**Median**	**Mean**	**Upper Quart.**	**Max.**
`w_p6` (*K1_AUF* & *KIGA_PANEL*)	17.522	52.397	79.559	105.899	125.470	3846.498
`w_p6` (*KIGA_IND*)	11.852	119.884	196.138	296.012	339.771	4501.028
`w_p6_joint`	3.123	37.049	56.396	77.996	91.182	2832.785

Acknowledgements This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Kindergarten, DOI:10.5157/NEPS:SC2:7.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.

References

Blossfeld, H.-P., Roßbach, H. G., & von Maurice, J. (Eds.). (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS) [Special Issue]: Zeitschrift für Erziehungswissenschaft* (Vol. 14). Wiesbaden: VS Verlag für Sozialwissenschaften.

Chu, A., Brick, J. M., & Kalton, G. (1999). Weights for combining surveys across time and space. In *Bulletin of the international statistical institute: 52nd session, contributed papers* (Vol. Book 2, p. 103-104).

Hellrung, M., Bockelmann, J., Schneider, C., Waschk, A., & Hillen, P. (2013). *Methodenbericht: NEPS Startkohorte 2 Haupterhebung – Frühjahr 2013 (A14, A14 A)* (Tech. Rep.). Hamburg: IEA Data Processing and Research Center (IEA DPC). (Part of NEPS data documentation)

Kiesl, H. (2016). Indirect sampling: A review of theory and recent applications. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *10*(4), 289-303. doi: 10.1007/s11943-016-0183-3

Statistisches Bundesamt. (2016). *Bildung und Kultur: Allgemeinbildende Schulen Schuljahr 2015/2016 (Fachserie 11, Reihe 1)*. Wiesbaden. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Schulen/AllgemeinbildendeSchulen2110100117004.pdf?__blob=publicationFile

Steinhauer, H. W., Aßmann, C., Zinn, S., Goßmann, S., & Rässler, S. (2015). Sampling and weighting cohort samples in institutional contexts. *AStA Wirtschafts- und Sozialstatistisches Archiv*, *9*(2), 131-157. doi: 10.1007/s11943-015-0162-0

Steinhauer, H. W., Zinn, S., Gaasch, C., & Goßmann, S. (2016). *NEPS Technical Report for Weighting: Weighting the sample of Kindergarten Children and Grade 1 Students of the National Educational Panel Study (Wave 1 to 3)* (Working Paper No. 66). Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_LXVI.pdf

Wolter, K. M. (2007). *Introduction to variance estimation* (2nd ed.). New York: Springer.

Würbach, A. (2018a). *Samples, Weights, and Nonresponse: the Kindergarten Cohort of the National Educational Panel Study (Wave 7)* (Technical Report). Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC2/7-0-0/SC2_7-0-0_W.pdf

Würbach, A. (2018b). *Samples, Weights, and Nonresponse: the Kindergarten Cohort of the National Educational Panel Study (Waves 1 to 6)* (Technical Report). Bamberg: Leibniz Institute for Educational Trajectories. Retrieved from https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC2/6-0-1/SC2_6-0-1_W.pdf

Zinn, S., Würbach, A., Steinhauer, H. W., & Hammon, A. (2018). *Attrition and selectivity of the neps starting cohorts: An overview of the past 8 years* (NEPS Survey Paper No. 35). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study.