Eduvest – Journal of Universal Studies
Volume 5 Number 6, June, 2025
p- ISSN 2775-3735- e-ISSN 2775-3727

Comparison of the Accuracy of Stratified Random Sampling and Simple Random Sampling Methods in National Assessment (AN)

Januar Pribadi[1], Achmad Ridwan[2], Awaluddin Tjalla[3]
Pusat Asesmen Pendidikan, Kemendikbudristek, Indonesia
Universitas Negeri Jakarta, Indonesia
Email: trijanuardi@gmail.com

ABSTRACT

Sampling methods are crucial for large-scale assessments. International surveys like PISA, TIMSS, and PIRLS use stratified random sampling (StRS) to enhance estimation accuracy, ensure representation of all subpopulations, and provide efficient administration. Similarly, Indonesia's National Assessment (AN) applies StRS, dividing populations by school size, class size, and gender. However, the accuracy of the AN sampling method, including its reliability and validity, has not been tested since its 2021 implementation. This study compares the reliability and validity of the AN sampling method to simple random sampling (SRS). Reliability is assessed by the consistency of estimates across repeated sampling, indicated by small standard error (SE) and confidence intervals (CI). Validity measures how accurately sample estimates reflect population parameters, evaluated through Mean Square Error (MSE). Using AN data from 1.9 million junior high school students out of 4.2 million, the analysis shows no significant differences in national population parameters between StRS and SRS. Both methods produce similar mean estimates (55) and standard deviations (10.7). However, StRS demonstrates greater variability in weights, reflecting its ability to account for sampling structure. At the school level, StRS outperforms SRS, yielding narrower CI and MSE ranges, highlighting its superior reliability. While MSE differences are statistically significant, their practical impact is minor due to the small effect size and large dataset. These results suggest StRS is more reliable for school-level reporting.

KEYWORDS *standard error, mean square error, national assessment, simple random sampling, stratified random sampling*

INTRODUCTION

Sampling methods play an important role in large-scale assessments. International surveys such as PISA, TIMMS, and PIRLS use sampling methods *stratified random sampling* (StRS) . Similar to international surveys, the National Assessment (AN) also uses the sampling method (OECD, PISA 2022 Technical Report, 2024; Almaskut, LaRoche, & Foy, 2023; Mang, Küchenhoff, Meinck, & Prenzel, 2021; LaRoche & Foy, 2020) *stratified random sampling* (Pusmendik, 2024a). AN is a form of evaluation of the education system by the Ministry of Education, Culture, Research, and Technology (Kemendikbudristek) at the primary and secondary education levels in formal and non-formal channels, the AN component consists of Minimum Competency Assessment (AKM), Character Survey, and Learning Environment Survey. AN participants included all principals, all teachers, and student representatives as many as 30 students for grade 5 and 45 students for grades 8 and 11 from each school throughout Indonesia who were randomly selected (Machromah et al., 2021; Megawati & Sutarto, 2021; Widarti et al., 2022).

Sampling method *stratified random sampling* (StRS) is one of the probability sampling methods where the population is divided into several subgroups called strata, then in each strata the sample is as much *n* units selected by the *simple random sampling* (Lohr, 2022; Wu & Thompson, 2020; Taherdoost, 2016; Levy & Lemeshow, 2013). This method has several advantages, namely A higher level of estimation accuracy Well compared to other methods, all recognized subpopulations are represented in the sample, the ability to calculate the accuracy level of the subpopulation measurements, efficient administration as it allows for small

sampling or different sample administration for each strata. Then what is meant by the *simple random sampling* (SRS) is a combination of the appearance of a number of samples (*n*) of the population (*N*) has the same probability (Lohr, 2022; Wu & Thompson, 2020; Taherdoost, 2016; Levy & Lemeshow, 2013). The advantage of SRS is that its sampling process is simple and every member of the population has an equal chance of being selected. (Berndt, 2020)In AN the strata of school size, class size, and gender are used. What is meant by school size is the number of students at the AN sample level (grades 5, 8, and 11). Then what is meant by class size is the number of students in each class. And gender is the number of males and females in each class (Pusmendik, 2025)

Since it was first implemented in 2021, the sampling method at AN has not been tested for accuracy. This is important Because the concept of accuracy has to do with reliability and validity. The AN sampling method is tested by comparing reliability and validity with other sampling methods, in this case SRS. In addition, there are two other reasons why the sampling method in AN needs to be tested for accuracy. The first is, based on the (Lohr, 2022). The monitoring report on the implementation of AN from the Center for Education Assessment (Pusmendik), some schools feel disadvantaged because the students selected as samples are students who have low ability, so they are considered not representative of the school. The second reason is that the proportion of the sample to the population in the AN is relatively large compared to the PISA international study. So it needs to be confirmed whether the large proportion of AN is enough to use SRS or still have to use StRS. Based on data from the Pusmendik, in AN 2022 the proportion of samples – the population at the junior high school level is 44.74% which is obtained from dividing the number of student samples as many as (Pusmendik, 2024b)1,908,413 with a total population of 4,262,615. Meanwhile, in PISA 2022, the proportion of Indonesian samples was only 0.04% (13,439 samples out of a population of 3,790,846 students). In PISA 2022, the average proportion of the sample – population in PISA participating countries was 16% with the smallest proportion belonging to the United States, which was only 0.01% (4,552 samples from a population of 3,661,328 students) and the largest proportion was Macau with a proportion of 99% (4,384 samples from a population of 4,423 students).

Then related to accuracy consisting of reliability and validity, this study defines the reliability of sampling methods as The ability to produce the same estimation of population parameters if the sampling process is carried out repeatedly, good consistency is demonstrated by *standard error (SE)* small ones. Taking into account that the actual population parameters are unknown and assuming the distribution of observation scores follows a normal curve, Levy & Lemeshow provide a formula for the confidence level of sampling with correction if the population number is known (Levy & Lemeshow, 2013; Lohr, 2022) *FPC/finite population correction*) as follows:

$$x' \pm z_{1-\left(\frac{\alpha}{2}\right)}(N)\sqrt{\frac{N-n}{n}}\left(\frac{SD_x}{\sqrt{n}}\right)$$

Where *x'* is an estimate of population parameters, *N* is the sum of the population, *n* is the number of samples, and *SDx* is the standard deviation, and is the percentile of the normal standard distribution. This study used an alpha (α) value of 5% in calculating $z_{1-\left(\frac{\alpha}{2}\right)}$*ONE*.

Validity is defined as How far is the approximate accuracy of the parameter estimation relative to the actual parameter value. The validity of the sampling method was measured using (Levy & Lemeshow, 2013; Lohr, 2022) *Mean Square Error* (*MSE*). The formula for calculating MSE is as follows: (Lohr, 2022; James, Witten, Hastie, & Tibshirani, 2013)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{f}(y_i))^2$$

Where MSE is the average square of error, n is the number of samples, is the score of the observation of subject $x_i i$, and is the prediction result of the score of subject $\hat{f}(y_i)i$. This study uses the normal distribution reference as a comparison of the distribution of observations.

Some of the research relevant to this study is research from Taherdoost who explained six stages on how to choose the right sampling method in the sampling process. Taherdoost explained that there are six stages in the sampling process, namely describing the target population, choosing a sampling framework, choosing a sampling technique, determining sample size, collecting data, and measuring the response rate. Then Berndt made an article about the characteristics of probability and non-probability sampling methods and explained the advantages and disadvantages of the types of sampling methods in the group of probability sampling methods, such as the SRS method which has advantages that are easy to understand and every member of the population has an equal opportunity to be selected. The disadvantage of the SRS method is that it has to have a complete list of population members which is sometimes difficult to meet. Then the systematic sampling method has the advantage of being easy to do compared to SRS and can choose a relatively balanced sampling between members of the population, while the disadvantage of this method is that sometimes important characteristics of the population can be missed to be captured by the sampling algorithm. In StRS, the selected sample is more representative of the population so as to increase the external validity and generalizability of the research and this method has the disadvantage of not being able to use it for populations that do not have clear unique characteristics and can separate populations based on their characteristics. On the (2016) (2016) (2020)*cluster random sampling* has the advantage of being relatively economical and can be done when dealing with a large population while maintaining variability in the population. The disadvantage of this method is that the assumption that each cluster has the same characteristics sometimes simplifies the population and creates bias.

Furthermore, Lin's research on the effects of (2018)*ONE* Against bias in meta-analysis with a small sample shows that the results of the meta-analysis should be treated very carefully, the research variant should not be enforced as a true variant, and the meta-analysis should consider sampling errors to avoid biased results. Altman and Bland's writings make it clear that (2014a; 2014b) *ONE* is a way of a study to reduce uncertainty. Sampling error is a statistical focus because a study uses one group of people who are one of many possible groups of the population represented by that group. Therefore, interval confidentiality and hypothesis testing were used to capture the uncertainty of the study results. However, it is sometimes often forgotten that *ONE* only reflects the consistency of a study and does not represent the accuracy of the research. One way to measure the accuracy of a study is to evaluate the model used in

the study. A controversial statement regarding the evaluation of the model was made by Willmott and Matsuura stating (2005). *Mean Absolute Error* (MAE) is better compared to *Root Mean Square Error* (RMSE) is related to models that predict weather because RMSE as a function is influenced by the magnitude of the error, the distribution within the error, and the number of samples, while the MAE is only influenced by the magnitude of the error. This study compares the reliability and validity of the AN sampling method to simple random sampling (SRS). Reliability is assessed by the consistency of estimates across repeated sampling, indicated by small standard error (SE) and confidence intervals (CI). Validity measures how accurately sample estimates reflect population parameters, evaluated through Mean Square Error (MSE). The optimal RMSE is used to predict Gaussian Error and the maximum MAE is used to predict Laplacian error. Social studies usually use the assumption of normality (Gaussian) in testing hypotheses, while Laplacian distributions are more often used to describe distributions with sudden changes.

METHOD

This study uses a quantitative approach to answer research problems. Quantitative research is a research method approach that aims to test theory by examining the relationships between measurable variables so that they can be analyzed using statistical procedures (Creswell & Creswell, 2018). This study uses secondary data from the Pusmendik in the form of AKM numeracy scores at the junior high school level taken during the implementation of the 2022 AN. This study uses descriptive analysis techniques and comparative analysis techniques of two groups. In the descriptive analysis, a description of the central tendency of the research variables will be carried out. In the comparative analysis of the two groups, it will be compared *ONE* and *MSE* from the StRS and SRS methods.

In order for the two methods to be compared, the same sample data will be applied as StRS and SRS. Although the research data is the result of the StRS process, it can be applied as if it were SRS by estimating population parameters based on the SRS method. The following formula for calculating the total estimate, average, and population variance using the StRS method refers to the explanation of the Levy & Lemeshow and Lohr. In StRS, the sample weight () is the number of units in the population represented by the sample and formulated by $w_{hj}w_{hj} = N_h/n_h$, where is the number of populations in the strata $N_h h$ and is the number of samples in the strata $n_h h$. Estimation The total value () in StRS is calculated using the following formula:$\hat{t}_{strs}$

$$\hat{t}_{strs} = \sum_{h=1}^{H} N_h \, \bar{x}_h$$

Where N_h is the number of strata members and is the average for each strata. The estimated $\bar{x}_h$average value ($\bar{x}$) in StRS is calculated using the following formula:

$$\bar{x}_{strs} = \sum_{h=1}^{h} \frac{N_h}{N} \bar{x}_h$$

Where h is a member of the strata, N_h is the number of strata members, N is the total number of members of the population, and is the average of each strata$\bar{x}_h$. The estimated value of the variant () of StRS is calculated using the following formula:S^2_{strs}

$$S_{strs}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{x}_h)^2$$

Where *h* is strata, n_h is the number of samples in strata, is the number of strata members, is the result of observation of subject $N_h x_{hi} i$ in strata *h*, and is the average in strata $\bar{x}_h h$. *Standard deviation* (*SD)* can be calculated by squared the estimated value of population variance, so that $SD_{strs} = \sqrt{S_{strs}^2}$.

Furthermore, for the calculation of total estimates, averages, and variances the SRS method refers to the formulas described by Levy & Lemeshow and Lohr. Sample weight on SRS (*WSRS*) calculated using the formula *wsrs= N/n*. where *N* is the sum of the population and *n* is the number of samples. Estimated value total () in SRS uses the following formula:$\hat{t}_{srs}$

$$\hat{t}_{srs} = \sum_{i=1}^{N} x_i$$

Where is the result of observation of population members $x_i i$ and *N* is the number of population members. The estimated average value ()$\bar{x}_{srs}$ in the SRS is calculated using the following formula:

$$\bar{x}_{srs} = \frac{1}{n} \sum_{i \in sampel} x_i$$

Where *n* is the number of samples, *i* is the sample unit, and is the result of the measurement of the sample unit $x_i i$. The estimated variance in SRS () is calculated using the following formula:S_{srs}^2

$$S_{srs}^2 = \frac{1}{N - 1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

Where *n* is the number of samples, *N* is the number of populations, x_i is the result of sample *observation i* and is the average of the variable $\bar{x}x$. The standard *deviation* formula (*SDsrs*) in SRS is the same as in StRS, which is the square root of the population variance estimate, so . $SD_{srs} = \sqrt{S_{srs}^2}$

RESULTS AND DISCUSSION

Table 1. Central Tendencies, Weights and Population Estimates

Central Tendencies		**StRS**	**SRS**
	N	4.262.615	
	n	1.908.413	
	Rerata	2,23	2,23
Weight	*SD*	2,08	-
Sampling	Varian	4,34	-
	Range	1 – 48	-
	Rerata	55	55
Estimation	*SD*	10,7	10,7
Population	Varian	114	115
	Range	0 – 100	0 – 100
***CI* 5% Population**		0,0168	0,0169

***MSE* Population**	230	230

The results of the descriptive analysis showed that in terms of population, there was no difference in population parameters between the sampling results of StRS and SRS. Table one shows a population of 4.2 million students and a sample of 1.9 million students. In addition, table one also shows the central tendency of population weight and estimation where the average sampling weight for StRS and SRS is the same as 2.23. Only the SRS weight does not have *an SD* and *variant* because for SRS all samples have the same weight, all samples on SRS are considered to represent about 2.23 students. Meanwhile, StRS has a weight variant of 4.38 and SD a weight of 2.08 with the smallest weight range of 1 and the largest 48. In StRS, the sample that has a weight of 1 is likely to be a student whose entire student in the school is used as an sample, while the student with a sample weight of 48 is likely to be the only student selected from his class to be the AN sample and happens to be from a large class.

For the population parameters, as shown in table one, it can be seen that the population parameters between the StRS and SRS methods are not different. The average between the StRS and SRS methods is both 55, with an SD of 10.7. As for the variants between the StRS and SRS methods, there are slight differences, namely 114 and 115. The reliability (*CI)* and validity between the StRS and SRS methods are also not much different. For the confidence level (*CI*) at 5% is around 0.02 for both the StRS and SRS methods and MSE is 230 points, both for the StRS and SRS methods.

The results of the estimation of population parameters between the StRS and SRS methods which are very similar are likely due to the fact that the data used is sampling data from the StRS method, the difference between the StRS and SRS methods in this study only exists in the calculation of sample weighting. In simulation studies where the StRS and SRS methods were compared in their entirety, it was shown that the StRS method was better than SRS. As researched by Ding, et al. The experiment of estimating electricity consumption in circuits was studied by experimental methods, where sampling was carried out with a measure of electrical power when the circuit was electrified showed that the StRS method was much more reliable than the SRS method because it had a much smaller error. Then simulation research from Wibowo, et al. showed that the StRS method has *HERSELF* and *MSE* which is better compared to the SRS method. The research is a simulation research using data from the 2011 PODES survey. The difference in results between similar studies can be explored by conducting a simulation study based on AN empirical data.

Furthermore, a school-level analysis is carried out, where the reliability (*THERE*) and validity (*MSE*) between the StRS and SRS methods compared at the school level. This analysis is important to do because the results of the AN are reported to the public in the form of education report cards at the school level. To investigate the accuracy of sampling results at the school level, a calculation will be carried out (Mendikbudristek, 2022). *THERE* and *MSE* in each school that participates in AN. Based on data, there are 58,292 junior high schools participating in AN 2022 with 221 schools that only have 1 student. These schools were not included in the analysis at the school level because they did not have a variance in scores, so the analysis at the school level involved 58,071 AN schools. At the school level, using the StRS range method *CI* is 0 – 17.5 and the MSE range is 0.02 – 1.641. While using SRS, the range *CI* is 0.14 – 34.5 and the MSE range is 0.008 – 1.387.

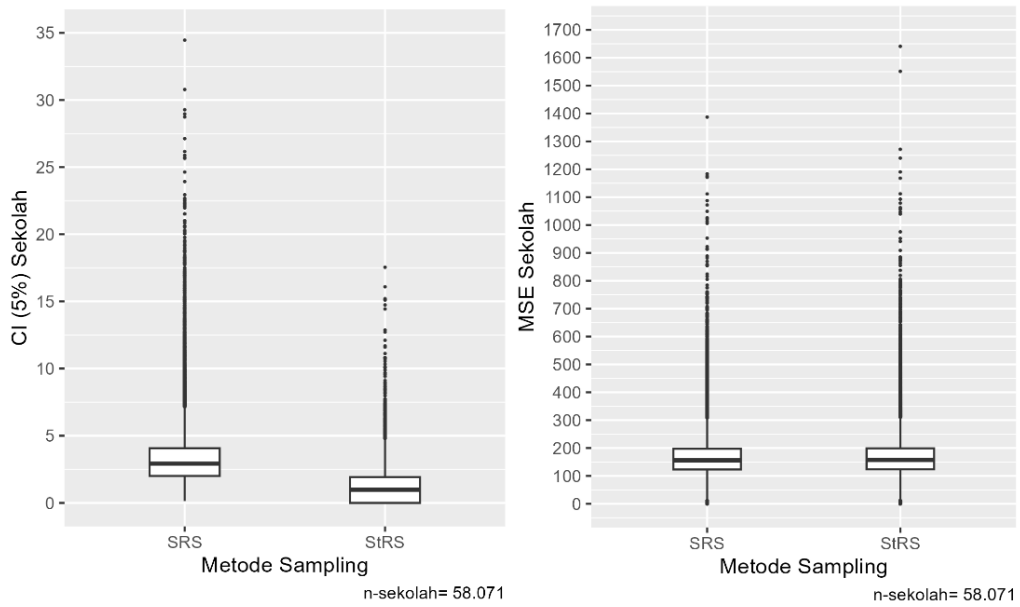

Figure 1. *Boxplot* CI 5% and *school MSE* between StRS and SRS methods

Figure 1 shows *boxplot* Spread *CI* and *MSE* at AN school. In the picture it can be seen that in general *CI* The SRS method is spread at higher values and has a much wider range compared to the spread *CI* StRS. This is also proven by the test *Wilcoxon Signed-Rank* pairing between *CI* StRS and SRS methods of $W = 169.219.219$, $p < 0.01$ with a relatively large effect size, i.e. $r = 0.70$. A more detailed observation of the (Gignac & Szodorai, 2016)*CI* at the school level, it shows that the StRS method has better reliability than the SRS method when all students in one school are taken as samples. In schools where all students are taken as samples, they have a high score $CI = 0$, whereas using the StRS method, the CI value ranges from 0 – 0.14. This happens because StRS in AN takes into account schools as strata so that it recognizes the number of students in each school. Meanwhile, the SRS method only recognizes the number of student populations. From the analysis at the school level, it can be concluded that the StRS method can better distinguish the school average from the distribution *CI* which is smaller compared to the SRS method.

Next, the comparison *MSE* between the StRS and SRS methods also showed significant differences with $W = 943.859.263$, $p< 0.01$. *Despite MSE* These two methods proved to be significantly different, but with a small effect size, namely $r = 0.10$, this difference is likely due to the large amount of data. despite the distribution *MSE* The StRS and SRS methods differ significantly with small effects, so the significant differences are less meaningful. Masha & Vetter (2018)explains the larger the sample count, the smaller the size of the effect produced, this is related to the power of drawing conclusions to avoid *error* type II (false negative). In the context of this study, the amount of data (n) 58,071 schools are relatively large, so it is very easy to be significant and produce a small effect size. The analysis of the difference between the two groups using the test *Wilcoxon Signed-Rank* pairs because the normality test on the data shows that the data is not distributed normally.

CONCLUSION

In the context of AN, the StRS and SRS methods show equal reliability and validity in population parameters, although it is necessary to emphasize the difference in sampling methods in this study only at the level of calculating the sampling weight and estimating population parameters (mean, SD, and variant). For further research, simulation research can be carried out because simulation studies help to evaluate statistical methods in controlled conditions such as in a laboratory room and allow to compare different conditions more economically (Boulesteix et al., 2020). Meanwhile, at the school level, the StRS method has been proven to have better reliability compared to the SRS method. It can be said that the StRS method is more reliable for describing subpopulations than the SRS method. On a practical level, these findings can be a guide for analyses related to AN involving comparisons between sub-populations, so the results will be more reliable if using the weight of the StRS calculation. Meanwhile, the validity comparison shows that both the StRS and SRS methods have a relatively similar level of validity.

REFERENCES

Abrahamowicz, M., Binder, H., Briel, M., Hornung, R., Morris, T. P., Rahnenführer, J., Sauerbrei, W., Groenwold, R. H. H., & Boulesteix, A.-L. (2020). Introduction to statistical simulations in health research. *BMJ Open, 10*(12), e039921. https://doi.org/10.1136/bmjopen-2020-039921

Almaskut, A., LaRoche, S., & Foy, P. (2023). Chapter 3: Sample design in PIRLS 2021. In M. v. Davier, I. V. Mullis, B. Fishbein, & P. Foy (Eds.), *Methods and procedures: PIRLS 2021 technical report*. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb9560

Altman, D. G., & Bland, J. M. (2014a). Uncertainty and sampling error. *BMJ*, g7064. https://doi.org/10.1136/bmj.g7064

Altman, D. G., & Bland, J. M. (2014b). Uncertainty beyond sampling error. *BMJ*, g7065. https://doi.org/10.1136/bmj.g7065

Berndt, A. E. (2020). Sampling methods. *Journal of Human Lactation*, 36(1), 1–3. https://doi.org/10.1177/0890334420906850

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Ding, C.-S., Haieh, C.-T., Wu, Q., & Pedram, M. (1996). Stratified random sampling for power estimation. In *Proceedings of International Conference on Computer Aided Design* (pp. 576–582). IEEE. https://doi.org/10.1109/ICCAD.1996.569913

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development, 15*, 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning with applications in R* (2nd ed.). Springer.

Kepala-BSKAP. (2024). *Keputusan BSKAP Kemendikbudristek No. 019/H/KP/2024 tentang Pedoman Penyelenggaraan AN*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

LaRoche, S., & Foy, P. (2020). Chapter 9: Sample implementation in TIMSS 2019. In M. O. Martin, M. v. Davier, & I. V. Mullis (Eds.), *Methods and procedures: TIMSS 2019 technical report*. TIMSS & PIRLS International Study Center, Boston College.

Levy, P. S., & Lemeshow, S. (2013). *Sampling of populations: Methods and applications* (4th ed.). John Wiley & Sons.

Lin, L. (2018). Bias caused by sampling error in meta-analysis with small sample sizes. *PLOS ONE*, *13*(9), e0204056. https://doi.org/10.1371/journal.pone.0204056

Lohr, S. L. (2022). *Sampling: Design and analysis* (3rd ed.). CRC Press.

Machromah, I. U., Utami, N. S., Setyaningsih, R., Mardhiyana, D., Wahyu, L., & Fatmawati, S. (2021). Minimum competency assessment: Designing tasks to support students' numeracy. *Turkish Journal of Computer and Mathematics Education, 12*(14), 5480–5487.

Mang, J., Küchenhoff, H., Meinck, S., & Prenzel, M. (2021). Sampling weights in multilevel modelling: An investigation using PISA sampling structures. *Large-scale Assessments in Education, 9*(1), 1–39. https://doi.org/10.1186/s40536-021-00099-0

Mascha, E. J., & Vetter, T. R. (2018). Significance, errors, power, and sample size: The blocking and tackling of statistics. *Anesthesia & Analgesia, 126*(2), 691–698. https://doi.org/10.1213/ANE.0000000000002741

Megawati, L. A., & Sutarto, H. (2021). Analysis numeracy literacy skills in terms of standardized math problem on a minimum competency assessment. *Unnes Journal of Mathematics Education, 10*(2), 128–135. https://doi.org/10.15294/ujme.v10i2.49540

Mendikbudristek. (2021). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi No. 71 Tahun 2021 tentang Asesmen Nasional*. Kemendikbudristek.

Mendikbudristek. (2022). *Permendikbudristek No. 9 Tahun 2022 tentang Evaluasi Sistem Pendidikan oleh Pemerintah Pusat dan Pemerintah Daerah terhadap PAUD, Dikdas, Dikmen*. Kemendikbudristek.

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

OECD. (2024). *PISA 2022 technical report*. OECD Publishing. https://www.oecd.org/en/publications/pisa-2022-technical-report_01820d6d-en.html

Pusmendik. (2024a, November 12). *Simulasi AKM*. https://pusmendik.kemdikbud.go.id/an/simulasi_akm

Pusmendik. (2024b). *Laporan Monitoring AN 2024*. Pusat Asesmen Pendidikan.

Pusmendik. (2025, January 8). *FAQ Asesmen Nasional*. https://pusatinformasi.raporpendidikan.kemdikbud.go.id/hc/en-us/articles/38597276705305

Salkind, N. J. (2006). *Encyclopedia of measurement and statistics*. SAGE Publications.

Taherdoost, H. (2016). Sampling methods in research methodology: How to choose a sampling technique for research. *International Journal of Academic Research in Management, 5*(2), 18–27.

Widarti, H. R., Rokhim, D. A., Septiani, M. O., & Dzikrulloh, M. H. A. (2022). Identification of science teacher practices and barriers in preparation of minimum competency assessment in the Covid-19 pandemic era. *Orbital: The Electronic Journal of Chemistry, 14*(1), 47–56. https://doi.org/10.17807/orbital.v14i1.1695

Wibowo, A., Indahwati, Sumertajaya, I. M., & Astuti, E. T. (2015). Accuracy comparison of simple, systematic, and stratified random sampling for estimating population (Minimarket case in Indonesia). In *Proceedings of International Conference on Research, Implementation and Education of Mathematics and Sciences* (pp. 168–175). Yogyakarta State University.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research, 30*, 79–82.

Wu, C., & Thompson, M. E. (2020). *Sampling theory and practice*. Springer. https://doi.org/10.1007/978-3-030-44246-0