**PAREZJA** 2/2022 (18) **STUDIA I ESEJE**
ISSN 2353-7914 DOI: 10.15290/PAREZJA.2022.18.02

Sławomir Pasikowski
Uniwersytet Łódzki
E-MAIL: slawomir.pasikowski@now.uni.lodz.pl ORCID: 0000-0002-0768-1596

# Measurement Uncertainty – its Place in Education Studies

## Introduction

The issue of uncertainty is not a popular one in educational research. It is rarely the subject of such research and even more rarely is it considered as a component of the observation process in pedagogical cognition. If it is addressed, it is usually as a psychological process, or as a threat to the achievement of goals, for example: a sense of uncertainty in the process of selecting learning content. For this reason, uncertainty is more often identified in educational research with decision theory and with risk theory and referred to psychological, economic, or political issues (Baranowski, 2017; Sahlberg & Senior, 2023). Uncertainty, however, is associated with such research much more frequently and strongly. Where information acquisition takes place under conditions of ignorance, lack of constancy, opacity of causes and consequences, technical limitations, where not all values of the observed variables are known and where inference cannot be reliable, there uncertainty is involved in the process of cognition and the use of its results. Hence, in the implementation of observational studies, uncertainty constantly accompanies both the collection of data and its processing and the making of predictions based on them. It cannot be eliminated. At most, it is possible to determine its magnitude and, possibly on this basis, attempt to reduce its scope. In empirical research methodology, the name uncertainty occurs in the context of randomness of events and in the context of the issue of measurement. A random event is the result of an experience that has not yet materialised. The result of a test measurement, the number of people in a group, or the feeling of satisfaction with a study are random variables that take on a certain value, but this value is uncertain. In other words, it is not known before the experiment is carried out. This does not preclude knowing the distribution of the values of a random variable. Its realisations can be predicted, thereby determining their uncertainty. Formulating conjectures about what is uncertain can be considered an attempt to measure uncertainty and uncertainty itself – a specific object of statistical thinking (Edge, 2019; Ostasiewicz, 2012). Effective instruments for conceptualising and measuring uncertainty are provided by the probability calculus, which is a mathematical theory. It enables uncertainty to be strictly quantified. For this reason, it is used in measurement theory to determine the results of a measurement, to estimate measurement errors, to assess the measurement accuracy and measurement precision, the accuracy and reliability of measurement tools, as well as to predict measurement results. In educational studies, measurement, generally understood as the creation of a representation of a characteristic, is one of the basic processes of building knowledge about the object of cognition. This is the case regardless of whether the research is used to construct theory or to diagnose the *status quo* and bring about change in the area of social practice. Sometimes, because of references to mathematics, measurement is wrongly associated with a narrow area of research practice, or a separate category of research. The aim of this article is to weaken this association, but first and foremost to give an idea of the uncertainty of measurement. I will also try to explain how this uncertainty can be dealt with in order to make it manageable. This is all the more important because measurement itself can be treated as a source of a judgments' certainty and a space for certain adjudication.

## Measurement

Measurement is a set of operations aimed at determining, from a given set of symbols, one that corresponds to a given value of the measured quantity. This quantity is called a measurand in measurement theory and is taken to be a feature of some object that can be qualitatively distinguished and quantified. Making a measurement requires the creation of a measurand model. In order to build this model, a quantity must be defined and then a system of reference values must be selected for it (e.g. Berka, 1983). This system is a set of symbols together with rules defining the relationships between them. These symbols do not have to be digits denoting numbers. Selecting such a set involves assigning symbols to different values of a measurand, but in such a way that the relations between them also map to the relations between values of that quantity. This activity is called scaling, and the mapping of a quantity by a symbol is called a scaling function. The function should be differentiable and unambiguous (bijection: injective and surjective function simultaneously). Through scaling, it becomes possible to map a quantity and the relationships between its values. The theoretical basis for mappings is the theory of functions, which provides models and a conceptual apparatus for describing relationships between sets: the set of values of a quantity, the set of symbols constituting the reference system of that quantity, the set of objects characterised by that quantity.

Measurement involves assigning a symbol to a given true measurand value, i.e. a realised value (random variable realisation: value of a random variable for a specific elementary event). For this, it is necessary to have a real reference system available for use in the act of measurement. Such a system makes it possible to compare the measurand with some benchmark quantity, to match and express the true value of the measurand in units of that quantity. An example is a rating scale. In its values, a measurand, such as an opinion or attitude of a person, is expressed. However, assigning a symbol to the true value of the measurand may not be adequate. Then the value observed in the act of measurement will be different from this reference value (true or conventional quantity value; said 'conventional' when this value is known or said 'true' when this value is unknown). This difference is called measurement error (or true measurement error) and as such is the oldest measure of measurement accuracy. Measurement error consists of random error and systematic error. The former is due to the unpredictability, or indeterminacy, of the conditions for obtaining the observed measurement value. The latter from the presence of regular influences on that value present in the act of measurement, such as a constant distortion of the measuring instrument, or the repeated misuse of that instrument. It is assumed in measurement theory that the true value is unknown. Consequently, the concept of true measurement error has been replaced by the concept of measurement uncertainty. Its evaluation in practice leads to the determination from a reference value system (e.g. a set of real numbers R) of not a single value, but a subset of values. This subset is called measurement result. The value of this subset is called measured value, or observation, observed in the act of measurement. This means that the measurement provides information about an interval of values covering the true value of the quantity that characterises the object to be measured for that quantity. The shorter this interval is, the more unambiguous the result of the measurement is, but thus less certain. A simple example would be the estimation of the interval in which a task is planned to be performed.

## Uncertainty of measurement

Measurement uncertainty is operationalised as a parameter of the dispersion of observed values that are attributed to the measurand (JCGM 200:

2012, 2008; JCGM 1004: 2009, 2009). Thus, in order to define it, it is necessary to determine the interval of values that will cover the unknown value of the measurand. This can be done in two ways (e.g. Bucher, 2012): method A ($u_A(x)$) for a series of measurements, and method B ($u_B(x)$) for a single measurement. In method B, it is necessary to assume a standard deviation for the probability distribution of the measurand, which can be based on knowledge of the measurement tool properties (e.g. SEM – standard error of measurement), previous measurements, or a well-established knowledge of the distribution of this measurand. The calculation of the uncertainty of method B is relatively straightforward, therefore method A is presented below. More information on method B can be found in metrology literature. In order to determine the measurement uncertainty for a series of observations, care should be taken at an early stage to correct for systematic errors (including gross errors) and then to determine the expected value estimator and the variance estimator based on these. These estimators for measurements made on interval and quotient scales are the arithmetic mean $\bar{x}$ and the standard deviation $s_x$. These are used to determine the next estimator, which is the standard deviation of the arithmetic mean $s_{\bar{x}}$:

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}}$$

where $s_x$ is the standard deviation, n is the sample size and $x_i$ is the value of the i-th observation.

The estimator $s_{\bar{x}}$ enables the difference between the mean value and the unknown true value to be assessed. This estimator is a measure of the standard measurement uncertainty $u_A(x)$. The procedure concludes with the determination of the expanded measurement uncertainty, which consists in determining a confidence interval based on the standard measurement uncertainty $u_A(x)$ and the coverage factor k. The latter takes a value corresponding to the number of standard deviations in the probability distribution of the standardised random variable, which indicates the confidence level at which the unknown true value of the measurand is inferred. The confidence level, is a probability taking the value $1 - \alpha$, where $\alpha$ is the assumed probability of an error in inference. Probability is essentially a measure of the random events uncertainty. For a probability of 0.99 (99%) and a probability of 0.95 (95%), the k-factor is $k = 3$ and $k = 2$, respectively.

$$U(x) = u_A(x)k$$

However, with a small number of measurements, e.g. with $n < 120$, the estimators may deviate greatly from the true value and the variance of the general population distribution of the measurement results. Therefore, in order to estimate the true value at the assumed confidence level, the so-called Student's t coefficient ($t_{\alpha, n-1}$) should be used instead of the k coefficient, which is the critical value of the Student's t distribution for a given number of degrees of freedom and an assumed probability α.

The above description can be illustrated by an example of five measurements made with the knowledge test X carried out on person O, in which the quantity measured is expressed by a number of points taking values from 0 to 35, in such a way that the more knowledge, the more points in the test. The following values were observed: $x_1 = 23$, $x_2 = 25$, $x_3 = 30$, $x_4 = 28$, $x_5 = 20$. The arithmetic mean of the measurements was $\bar{x} = 25.20$, the standard deviation $s_x = 3.96$, the standard uncertainty of measurement $u_A(x) = 1.77$, and at $t_{0.05,\ 5-1} = 2.57$ the expanded uncertainty of measurement $U(x) = 4.55$. Thus, the measurement result is $x = 25.20 \pm 4.55$ (test points), which should be interpreted to mean that, with a probability of 95%, the interval with ends 20.65 and 29.75 will cover the unknown true value of knowledge X held by person O.

In the example given, the issue of accuracy and reliability of the test has been ignored, assuming that the test has been corrected for systematic measurement errors. However, it would be necessary to determine a composite uncertainty of measurement ($u_C(x)$), which is the square root of the sum of the standard uncertainties, if the evaluation of the measurement uncertainty were also to take into account the uncertainty arising from the measurement method and from the person performing the measurement. The method uncertainty can then be determined from the standard error of measurement (SEM) and the standard error of estimation (SEE), either determined separately or read from the test specification (cf. Chadha, 2009; Hornowska, 2009). On the other hand, the uncertainty coming from the person performing the measurement could be determined by, for example, the number of points misread during a single act of measurement.

## Conclusions

Measurement plays a key role in observational studies, including those relating to education, by providing a representation of the value of the characteristic being observed. This makes it possible to compare observable events with the content and structure of sentences that are consequences of theories or models describing classes of these events. These representations can be numerical, but also linguistic. Measurement is a complex process of mapping intersubjectively, or intrasubjectively, observable objects. It is wrong to reduce it to an operation of collecting data or counting observations. While it does provide uncertain knowledge, it also makes it possible to determine the degree and extent of knowledge uncertainty. Without this, single observations or variable results would leave doubts and confusion rather than provide answers to the questions posed. Combining uncertain knowledge with knowledge about the extent of uncertainty leads to useful knowledge (Rao, 1994, p. 54) that can inform decision-making and new information, but also allows judgements to be made about the degree of unreliability of the consequences derived from it.

## BIBLIOGRAPHY

Baranowski, M. (2017). Education in times of uncertainty. Uncertainty in education. A critical approach. In: A. Odrowąż-Coates, & S. Goswami (Eds.), *Symbolic violence in socio-educational contexts. A post-colonial critique* (pp. 63–72). Wydawnictwo Akademii Pedagogiki Specjalnej im. Marii Grzegorzewskiej.

Berka, K. (1983). *Measurement. Its Concepts, Theories and Problems*. D. Reidel Publishing Company.

Bucher, J.L. (2012). *The Metrology Handbook*. ASQ Quality Press. Chadha, N.K. (2009). Applied Psychometry. SAGE.

Edge, M.D. (2019). *Statistical Thinking from Scratch. A Primer for Scientists*. Oxford University Press.

Hornowska, E. (2009). *Testy psychologiczne. Teoria i praktyka*. Scholar.

JCGM 104:2009 (2009). *Evaluation of measurement data – An introduction to the "Guide to the expression of uncertainty in measurement" and related documents*. Joint Committee for Guides in Metrology.

JCGM 200:2012 (2009). *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. Joint Committee for Guides in Metrology.

Ostasiewicz, W. (2012). *Myślenie statystyczne*. Wolters Kluwer.

Rao, C.R (1994). Statistics and Truth. Putting Chance to Work. World Scientific Publishing.

Sahlberg, P., Senior, C. (2023). Education in an Uncertain Future: The Promise of Equity. In: W.O. Lee, P. Brown, A.L. Goodwin, & A. Green (Eds.), *International Handbook on Education Development in Asia-Pacific*. Springer. https://doi.org/10.1007/978-981-16-2327-1_15-1

## SUMMARY

The article deals with the issue of uncertainty. The importance of this issue in education studies is indicated. Selected ideas and models from the foundations of measurement theory and probability theory are presented and related to selected circulating opinions, as well as an example of the determination of measurement uncertainty.

KEY WORDS: uncertainty, measurement, observational studies