

Hanna Komorowska
hannakomo@data.pl
SWPS University of Social Sciences and Humanities
ORCID: 0000-0002-9395-2734

The predicament of assessment

Abstract: James Clifford (1988) in his seminal book *The Predicament of Culture* analyses deceptive accounts of other cultures presented from one's own perspective. Drawing on this analysis I would like to reflect on examinations and certification looking at quantitative evaluation seen from the qualitative standpoint as well as at qualitative assessment viewed from the quantitative stance.

Keywords: language, evaluation, assessment, quantitative, qualitative

1. Introduction

In an attempt to analyse the concept and history of evaluation in order to identify ways of future development, we first need to select an approach to reflection which best serves the purpose. Three types of reflection have been distinguished by philosophers and researchers in the field of education on the basis of the purposes it may serve. These are: *technical reflection* engaged in deciding paths leading to goal achievement, *practical reflection* helpful in evaluating processes and *critical reflection* recommended for the examination of dilemmas and obstacles encountered in research and study (Habermas, 1971; 1984; Hatton & Smith, 1995; Van Manen, 1995).

Reflection in its strong form calls for suspension of all assumptions as required in philosophy by the phenomenological concept of *epoché* (Husserl,

1913/1982) or later in sociology by the notion of *bracketing* (Garfinkel, 1968). Weak forms based on the ability to take perspectives other than one's own, such as Piaget's and Selman's *decentring* (Piaget & Inhelder, 1936/1969; Selman, 1980), later followed in FLT/SLA by Byram (2008) seem to be more realistic and, therefore suitable for the analysis of evaluation in language education. Flexibility in perspective-taking is indispensable today when, after periods of measurement, description and judgment, the fourth generation of evaluation has been fully installed with its contextual, dialogic, interactive, responsive-constructivist approach to examinations and certification (Guba & Lincoln, 1989; 1994; Korporowicz, 2012).

Reflecting on evaluation, being the aim of the present text, calls for the clarification of both terms as neither is clear nor transparent. This may have some advantages as, according to François, the career of an approach or a concept is inextricably connected with opaqueness and metaphoricity of its meaning as well as with a vast array of misunderstandings it breeds (François, 2010: 27).

Evaluation poses numerous problems related to aims and processes, these, however, cannot be solved without agreeing on terminology used, since chaos in this field means traps, pitfalls and stumbling blocks for both examiners and examinees. Although in many publications the terms *evaluation* and *assessment* are used synonymously, assessment is usually considered to be a concept broader than evaluation. If no difference in scope is believed to be found between the two, the term *evaluation* is more often used to describe formal, objective and standardized procedures, while *assessment* – informal, subjective and non-standardized ones. Misunderstandings also occur due to a non-uniform usage of the term *certification*, as sometimes it means no more than issuing a certificate of attendance, sometimes of course completion, sometimes of a certain kind of achievement, while on other occasions it might mean fulfilling official requirements needed to obtain professional qualifications for a future job.

Part of the terminological problems result from the fact that purposes of evaluation differ considerably. Quite often the aim of evaluation is *selection* carried out in order to choose several individuals from a larger group, e.g. when there is a limit set on a number of candidates. This way of using evaluation procedures is probably the oldest and dates back to imperial examinations in China which started twenty-four centuries ago and became fully formalized in the 7th century (Elman, 1991; Shichuan, 2015). Sometimes there is no set limit, but the aim is *deselection* of applicants who do not meet the requirements, e.g. when their knowledge of a selected subject or a group of subjects

is deemed insufficient. Most probably the first examinations structured for this purpose were the 19th century Indian Civil Service Exams (Bachman, 2015). In certain situations, evaluation procedures are needed to show *differentiation* in a group, e.g. when some kind of prioritizing is the aim; the Polish secondary school leaving *matura* exams have served this purpose since the moment they replaced college entrance examinations. It should, however, be noted that the introduction of an arbitrary cut-off point may add a deselection function to the prioritization one. In other contexts, *certification* confirms that certain proficiency standards have been met. One of the oldest examples of certification in higher education is the doctoral diploma in canon law granted to Nicolaus Copernicus at Ferrara, a reputed *refugium pauperorum*, where students unable to afford certification at more expensive Italian universities used to take their examinations. A certificate or a diploma can function as a document *granting access to a profession or to higher levels* of education. Finally, evaluation might be geared toward an assessment of examinees' proficiency levels. This particular aim, which at the same time allows test takers to identify their strengths and weaknesses, leads us to the issue of examinations and certification in language education as a *diagnosis* – is a function that lies at the heart of a large number of popular international language examinations.

2. Qualitative vs. quantitative approaches to evaluation in language education

Beginnings of evaluation in language education and the move from informal to more formalized procedures date back to the beginning of the 20th century and the *Cambridge Proficiency in English* examinations. The Grammar-Translation Method was still widely used, a phenomenon understandable in the years of consolidating mass education. Its emphasis on reading comprehension and writing skills was not conducive to the implementation of oral approaches propagated by the Reform Movement. As subjective, integrative and holistic tendencies excluding any objective measurement prevailed at that time, this method can be considered an early version of the *qualitative paradigm*. Essays and translations present in the evaluation process, so strongly criticized now, do not deserve the contempt of today's educators, as in the early days they were an effective measure to familiarize school learners, coming from diverse, often poor family backgrounds, with high culture and literary language registers unknown to them. What is more, outside the English-speaking countries, translation from

L2 into the mother tongue was extensively used, which – together with the popularity of essay writing – contributed to the promotion of literacy and more advanced linguistic skills, also in the first language of the students, a feature of great significance for largely illiterate communities.

Toward the end of the interwar period, product-oriented approaches grew in importance and behaviourism, which reached the status of the dominant school of thought, directed harsh criticism at all subjective and integrative approaches. At the same time proponents of structuralism stressed the insufficiency of practical language learning and the need to introduce more linguistic content at the cost of high culture. The quantitative paradigm entered the scene signalling the arrival of the audiolingual era with its postulates to focus on decontextualized parts of the whole, which would allow for the analysis of quantifiable products inviting formal, statistical approaches. Summative evaluation with objective language testing was born right after World War II and dominated the scene for more than three decades.

The qualitative paradigm in its modernized form started fighting back at the beginning of the 1970s. The internationally reputed sociologist, Polish-born Stanisław Andrzejewski, taken into Russian captivity after the September Campaign of 1939 and a fugitive from Katyn, publishing under the name of Stanislaw Andreski, was one of the first critics of the quantitative approach. In his seminal book *Social Sciences as Sorcery* (Andreski, 1972), he attacked 'the wide acceptance of the dogma that nothing can be worth knowing that cannot be counted' (1972: 118) and stated that although he fully appreciates the usefulness of quantification when it is more than an academic camouflage, he is arguing against 'the soul-destroying taboo against touching anything that cannot be quantified' (1972: 145).

In language education, supporters of the qualitative approach turned against some testing techniques such as gap filling, matching and 'cloze' due to their lack of authenticity. They also went on to criticize the low communicative value of testing caused by examiners' behaviour and the absence of exchanges typical of everyday interaction. They also pointed not only to the stress factor impacting on test reliability, but also to the negative washback effect of final examinations on the learning process.

With the proliferation of language tests more anti-quantitative arguments were presented. It was pointed out that language tests test what is easy to test and not what is worth testing. Claims that language tests have the potential to enhance the transfer of training, thanks to which all skills developed during

test-taking are expected to have an effect beyond the classroom, were considered unfounded. Harmful washback was also warned against as opponents of quantitative approaches to measuring educational achievement alleged that objective tests had two undesired effects: that students do not learn what we want them to learn and that they do not learn the way we want them to learn.

Another set of problems with objective testing arose because of common errors in scoring: the same aspect such as grammar or spelling was scored several times in several different items. Points were deducted for wrong answers, which discouraged students from entering interaction and taking risks, i.e. from communication. Proportions planned in test specifications were frequently changed by the introduction of extended scoring in which several points were assigned to one test item, usually more difficult but not necessarily more important than the others. Strong criticism was also directed at the uniform standard for all.

Yet, the position of evaluation in the form of objective language tests was stable due to its practicality, a feature that cannot be ignored in mass education. Content validity, reliability, and objectivity, crucial for examiners' professional safety, were other important success factors. Objective testing also proved indispensable for certification.

Attempts to keep advantages of test-based evaluation, but to introduce changes at the same time, were strengthened when the qualitative research paradigm turned against the quantitative past, thus developing Andreski's earlier ideas. The qualitative approach proved to be well rooted in ethnography, cultural anthropology, phenomenology, social psychology and humanistic psychology, therefore the scientific basis for applying it in other areas was solid. Consideration of the non-quantifiable, emphasis on the context and more focus on the process were trends postulated by supporters of the qualitative paradigm, who also saw the value of inductive approaches, less structured data and a degree of researchers' subjectivity (Bogdan & Biklen, 1982; Denzin & Lincoln, 2005; Jacobs & Farrell, 2001; Richards, 2009).

3. Summative vs. formative evaluation

The qualitative turn in sociology and psychology had an almost immediate effect on education, which due to the promotion of progressive education, focus on personal growth and the development of curriculum theory and classroom research was ready for radical changes (Bruner, 1960; Bloom et al., 1964; Delamont, 1976).