

ANNA CZURA

## Adolescent learner perceptions of foreign language assessment: Critical incident analysis

**ABSTRACT.** This article reports on a study that aimed to explore adolescent learner perceptions of language assessment experiences and their beliefs about what makes a useful, valuable and high-quality assessment. The data was collected by means of the critical incident technique (CIT) in which 25 lower secondary learners (aged 14–16) were asked to recall both positive and negative instances of assessment they had experienced in the course of their language education. The results reveal that the participants held generally positive views of language assessment and paid attention to such aspects of assessment as reliability, appreciation and authenticity. On the other hand, the importance of grades was prioritised over the learning value of assessment, which is characteristic of performance-oriented assessment environment.

**KEYWORDS:** language assessment, learner perceptions, adolescent learners, critical incident technique (CIT).

### 1. INTRODUCTION

Research on language assessment usually focuses on the effectiveness of different methods or approaches to measuring learners' proficiency as well as on maintaining validity, reliability and other properties of good assessment. There are also debates about the superiority of different types of assessment and the interplay between summative and formative assessment forms. Although in the subject literature we can find studies that explore teachers' beliefs about assessment, little is known about learner perceptions of and beliefs about language assessment they encounter. In the meantime, learner perceptions of their immediate learning context and the way they construct the reality they experience exert a powerful influence on their approaches to learning and studying (Entwistle 1991; Struyven, Dochy &

Janssens 2005). Therefore, gaining an in-depth understanding of language learning necessarily entails the study of learner perceptions concerning different facets of this process, of which classroom assessment is an inherent part. With this in mind, the present study sets out to explore adolescent learner attitudes to and perceptions of language assessment they have experienced. In line with the social turn in foreign language acquisition studies, the research presented here adopts the qualitative methodology and makes use of the critical incident technique (CIT) with the objective of eliciting both positive and negative assessment-related situations the learners encountered in the course of their education. The application of this elicitation technique as a basis of an individual oral interview allowed the learners to report on the experiences they personally perceived as particularly important, memorable and influential. The results of the study are hoped to contribute to our understanding of what learners perceive as useful, valuable and high-quality assessment.

## 2. LEARNER PERCEPTIONS OF ASSESSMENT

In this article, *assessment* is understood as a variety of methods of collecting data about the process and product of learning. Its scope should not be limited to *testing*, which refers to the type of assessment that usually emphasises the product of learning and is administered through neutral, decontextualized tasks. In light of these definitions, assessment is a broader term and may embrace both formative and summative approaches to measuring foreign language proficiency in many contexts and for multiple purposes.

In modern language teaching, there is a growing focus on ethical aspects of language assessment. As Bachman (1990: 279) observes: "tests are not developed and used in a value-free psychometric test-tube; they are virtually always intended to serve the needs of an educational system or of society at large." Consequently, while designing a measurement tool and adopting a new assessment approach, test designers and language teachers need to take into account not only the construct or content, but also the social implications of assessment. Pedagogical consequences of assessment are of the utmost importance: a measurement tool can be called valid only if it exerts a positive effect on teaching and learning (Messick 1989: 27). For this reason, before an assessment tool is implemented, a number of questions need to be answered: What is the aim of the assessment? Will all parties involved benefit from it? What are the consequences of the test on the society and on an individual? How will teaching and learning be affected? How will the measurement be perceived by the learners?

Since teachers spend a large portion of teaching on assessment-related activities, each classroom has its own classroom assessment environment (Stiggins & Conklin 1992: 80), which depends not only on the teacher's decisions concerning the aims of assessment, choice of elicitation tasks, forms of feedback and grading policies, but also on how learners understand, evaluate and respond to the assessment practices they encounter. The extent to which learners are willing to engage in studying is shaped by their perception of the usefulness, value and importance of the appointed assessment task. A study conducted by Lizzio and Wilson (2013) in a higher education setting indicates that students' engagement depends largely on their subjective perception of the motivational value of the task. Additionally, unless the learners perceive a newly implemented assessment method as useful and beneficial, it may not be regarded as a valid tool of verifying their progress and achievement. In a group of adolescents, a few learners refused to engage in portfolio assessment as they considered it to be an additional burden with no tangible gains (Czura 2013).

There is, therefore, a reciprocal relationship between learner perception of classroom assessment environment they are part of and their approaches to learning. As Cheng, Wu and Liu (2015: 2) observe, "The ways teachers communicate their expectations to students, and the ways they provide feedback on how well these expectations are being met, help students form concepts of what is important to learn and how good they are at learning it". Depending on how assessment is perceived, it can be associated with praise and pleasure or constitute a source of threat (Pollard & Triggs 2000). Accordingly, since language assessment affects the learning process and learner perceptions of themselves as learners, their beliefs and understanding of assessment practices cannot be overlooked in research on foreign language learning and teaching. The findings of studies that investigate learner views of assessment practices may support teachers in the process of designing assessment that both meets the curricular requirements and caters for the learners' needs.

Earlier studies on learners' outlook on assessment in general education have usually adopted quantitative methodology and made use of Dorman and Knightley's (2006) Perceptions of Assessment Tasks Inventory (PATI), which focuses on such properties of assessment as congruence with planned learning, authenticity, student consultation, transparency and diversity, or Alkharusi's (2011) Perceived Classroom Assessment Environment Scale. The latter instrument is a two-factor scale that, on the basis of learner perceptions, distinguishes between *learning-oriented assessment environment*, in which emphasis is placed on the mastery of content material and the learning value of assessment practices, and *performance-oriented assessment envi-*

*ronment*, which entails interests in scores and grading and may involve normative comparisons of learners' performances (Alkharusi 2015).

Recently, a few attempts have been made to explore learner perceptions of assessment by means of qualitative research methods. For instance, oral diaries were applied in a longitudinal study that aimed to explore Finnish test-takers' perceptions of a high-stakes language test (Huhta, Kalaja & Pitkänen-Huhta 2006). Two other studies conducted respectively in primary school (Carless & Lam 2014) and high school settings (Xiao & Carless 2013) were based on a combination of a draw-a-picture technique and an interview and set out to investigate the learners' views of English language assessment experience as well as the emotions these experiences evoked.

### 3. METHODS

This study is to explore how adolescent learners view language assessment they have experienced in the course of their education by means of CIT. Through the content analysis of the recalled critical incidents, this article aims to gain insights into learners' attitudes to and perceptions of language assessment as well as assessment practices used in their immediate learning context. Accordingly, the following research questions were formulated:

1. What assessment experiences do the adolescent participants perceive as particularly positive and negative?
2. What aspects of assessment are of utmost importance to the learners?
3. Is the investigated learning context more typical of learning-oriented or performance-oriented assessment environment?

To the author's knowledge, there exist no empirical studies in which CIT has been applied with the objective of eliciting learner perceptions of foreign language assessment; therefore, it is additionally intended to analyse the applicability of this data elicitation technique in studies on learner perceptions in the field of second language learning and teaching.

#### 3.1. Participants

The study involved 25 lower secondary school learners (aged 15–16) who lived and studied in a large city in the south of Poland. The sample consisted of 14 girls and 9 boys who attended the same school and had three English lessons a week. Although the participants were taught by different teachers, language instruction in the school was based on the same syllabus and

course-book series, and the teachers had to adhere to a common assessment system approved by the school authorities. About a third of the learners took part in different forms of private tuition outside of the public school system. Since the analysed sample was relatively small, it must be underlined here that the study findings should not be generalized to other schools and contexts.

The analysis of learner perceptions of assessment should take into account the educational context in which the research is placed. Although the national curriculum encourages the promotion of an action-oriented approach and formative assessment, Polish schools cultivate a rather traditional and textbook-centred approach to language teaching, which gives preference to summative modes of assessment over more formative approaches. Language teachers tend to overemphasise the knowledge of grammar structures and linguistic accuracy rather than fluency and communicative skills. Lower secondary school ends with an obligatory written final examination which can be taken at either A2 or A2+ level of proficiency. The A2 examination includes only closed-ended tasks that aim at eliciting listening and reading comprehension, knowledge of language functions, as well as lexis and grammar skills, whereas in the A2+ version the last element is replaced with an open-ended writing task. The results of this high-stakes examination play an important role in the upper secondary school enrolment procedure.

### 3.2. Data collection

Language learner beliefs and perceptions are typically collected by means of questionnaires; however, since nowadays more qualitative and contextual approaches are gaining in popularity in studies on language learning and teaching, the critical incident technique was adopted in the present study. Chell (1998: 56) defines CIT as:

A qualitative interview procedure which facilitates the investigation of significant occurrences (events, incidents, processes, or issues) identified by the respondent, the way they are managed, and the outcomes in terms of perceived effects. The objective is to gain understanding of the incident from the perspective of the individual, taking into account cognitive, affective, and behavioural elements.

Tripp (1993: 3) adds that “a critical incident is an interpretation of the significance of the event”; consequently, apart from eliciting purely narrative data, CIT also encourages the respondents to interpret the recalled situation, examine their own reactions and critically evaluate the meaning of the incident and its consequences.

Although CIT was introduced by Flanagan (1954) in the 1950s and has been present in the social sciences ever since, it has not been adopted in research on foreign language learning and teaching until recently. The emergence of studies that make use of critical incidents to investigate different facets of language acquisition can be traced back to the turn of the 20th and 21st centuries and coincides with a growing interest in narrative research methods. Nowadays, CIT is applied in studies concerning sociolinguistic competence (Kuchuk 2012), intercultural competence (McAllister et al. 2006; McClure 2007) and learners' experiences of foreign language learning (Finch 2010). Gabryś-Barker (2012) and Farrell (2008), on the other hand, used critical incidents as a pedagogical tool designed specifically to generate pre-service FL teachers' reflection.

In the present study, the critical incidents were collected during oral interviews conducted individually in a quiet room or office. The participants were requested to recall and interpret one particularly positive and one particularly negative experience of having their language competences assessed at any level of education, both in and outside school. When needed, the critical incident was further defined as a particularly influential, memorable and important situation the learner wished/did not wish to be applied as a regular assessment practice. In order to elicit information concerning the learners' reactions, emotions and the employed coping strategies, the researcher asked a number of additional questions, for instance: How did you feel at the time? What was your reaction to the situation?

The interviews were recorded, transcribed and then content analysed by means of NVivo software. Participation in the study was voluntary, and all relevant parental permissions had been obtained before the interviews started.

#### 4. RESULTS

For the purposes of further analysis, a critical incident was defined as either a positive or negative, concrete and clearly identifiable situation in the past that involved assessment of language performance. Since some of the utterances offered a general evaluation of assessment and thus failed to meet these criteria, six positive and six negative accounts were discarded from further analysis. As a result, in total 19 positive and 19 negative reports were content analysed.

After the initial analysis, it was decided to code both positive and negative incidents in reference to the following main categories, which, as can be seen, are largely congruent with the principles of good assessment:

- face validity - entails learners' perception whether and to what extent the assessment measures the content and competencies it is supposed to measure,
- reliability - refers to the consistency of results, the overall fairness of assessment and the impact of external factors on the results of assessment,
- appreciation - involves the learners' feeling of their efforts being acknowledged and valued by the teacher,
- practicality - means that assessment stays within acceptable time constraints and does not involve excessive workload,
- authenticity - pertains to the extent to which the assessment involves real-life foreign language use, is interesting for the learners and allows them to work on their own initiative.

When appropriate, the respondents' utterances were coded into maximum two categories.

#### 4.1. The analysis of positive critical incidents

Figure 1 illustrates the aspects of assessment that the participants referred to in their accounts of positive critical incidents.

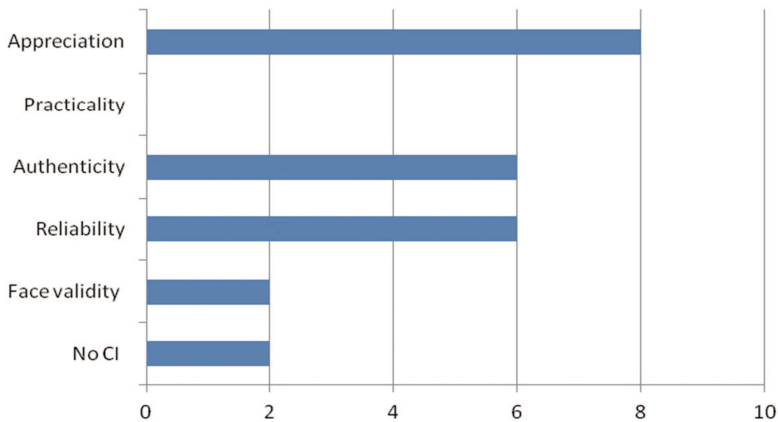


Figure 1. Aspects of assessment referred to in the positive critical incidents

As can be observed, the most frequently recalled category referred to the fact of being appreciated by the teacher. Five learners singled out the fact of being awarded a good grade for the effort they had made to prepare for

a test or a project as positive critical incidents. As one of such learners commented on receiving the highest grade for participating in a local language contest: "It was so nice. I succeeded and it was acknowledged by the teacher" (L18). It is worth underlining that for two respondents within this category receiving a good grade was not necessary – a verbal praise given by the teacher was perceived as a sufficient token of appreciation:

Learner: During a lesson the teacher asked a question and when I answered the teacher praised me.

Researcher: Did you get any grade?

L: No, although we get some pluses for good answers. It's normal in our class that the teacher praises us when we do something good.

R: So even though you didn't get a grade or a plus then, it was OK with you?

L: Yes. (L14)

Six accounts touched upon authenticity and reliability. Starting from the former, the learners seemed to have appreciated the forms of assessment that enabled them to use English in communicative and meaningful contexts. In one of such incidents, a learner recalled a group presentation assigned after a visit to the zoo: "It was great as we could write, present and say something in English. I like when we use English, not when we sit and do exercises" (L1). Project-based assessment was also praised by a few other learners, who enjoyed the possibility of working independently and on their own initiative. Another positive experience occurred during a school exchange in Spain when the learners were awarded a grade for their communicative skills: "The teacher observed how we used English and graded us depending on our competence, on whether we were really able to communicate or whether we simply used body language. It was quite cool" (L10).

In the utterances falling under the reliability category, learners focused on such aspects as the teacher's fairness, consistency of assessment and their own awareness of criteria applied in both teacher-based and peer assessment. The respondents' reflections on the critical incidents suggest that by and large they trusted their teachers' judgements regarding their performance:

The teachers always assesses as in a fair and just way. She's really competent in English and in assessment. (L2)

I think that, in general, our grades reflect what we have done (...). It's like that in the case of each test and each writing assignment. (L9)

Interestingly, in the account of one critical incident, assessment criteria were not presented or explained in any way; still, the learners felt that the grade they had been awarded truly reflected their performance.



As regards the incidents referring to validity, one learner appreciated that in a project-based assessment the teacher had paid more attention to language performance than to the drawing skills, which he considered his Achilles' heel. The second account focused on the usefulness of a test that had offered comprehensive and informative feedback: "In that test a student can get, e.g., 60% in grammar, 60% in vocabulary, 60% in writing, as we always have writing, and 60% in speaking. It's great as then we know what we suck at and what we need to do to get better" (L2). Finally, there were two learners who, despite holding a generally positive view of the assessment they experienced, were not able to single out any specific positive critical incident.

#### 4.2. The analysis of negative critical incidents

As illustrated in Figure 2, with regard to the negative critical incidents, situations describing different instances of violating the reliability category were recalled most often. The learners pointed out different cases of bias and favouritism as well as of subjective, unfair or inconsistent grading. As one of the respondents noted: "Once after I was absent, I took a makeup test and the tasks in that test were completely different from the ones my classmates had been given. It wasn't a different version of the test, it was a completely different test" (L13). As such a test is hardly likely to give results comparable to those of other learners, the respondent's negative evaluation seems justifiable in this situation. What is more, six out of ten learners supported their accounts of negative incidents with just and reasonable explanations that

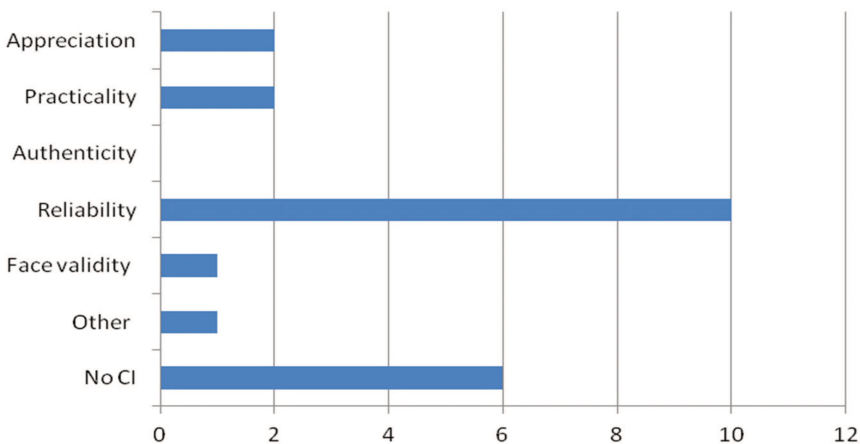


Figure 2. Aspects of assessment referred to in the negative critical incidents

suggested the violation of the reliability criterion in one way or another by the teacher. It is worth mentioning that while recalling the negative incidents, three learners added: “generally, assessment is very fair here” (L13), “in this school the teachers are fair”, and “otherwise, the way we are assessed is OK” (L10), which may indicate that such incidents are not regarded as commonplace in that school. Similarly, five out of six learners who failed to provide a negative incident claimed that they were generally pleased with the way they are assessed in English lessons. One of the respondents noted: “I can’t remember such a situation in English lessons. In the case of other subjects, yes; in the case of English – never” (L18).

Two learners referred to a situation in which, in their opinion, their efforts were not sufficiently appreciated by the teacher. The following fragment was coded into both appreciation and reliability category:

L: When I wrote one of the longest written assignments and I got one of the worst grades.

R: Why did you get a low grade?

L: I had lots of language mistakes, typos and so on. (L20)

It is clear there that the learner’s failure to accept the low grade stemmed from her unawareness of the assessment criteria, which, as she later added, had not been presented or discussed with the group at any point.

Finally, while commenting on a negative critical incident, one of the respondents noted that he dislikes:

when a teacher compares you to others, when you can’t do something. The teacher doesn’t adjust the scale to you. Someone is very good at something and another person isn’t. And both of them are assessed in the same way. In my opinion, it is wrong. (L13)

Whereas the feeling of injustice resulting from being compared to other learners may be understandable, the wish to have assessment criteria adjusted to the needs of each individual learner indicates a rather simplistic understanding of the role of classroom assessment and was, therefore, assigned to the ‘other’ category.

## 5. DISCUSSION

The present study aims to explore how English language assessment is perceived by adolescent learners in the selected context. Since the positive incidents clearly outnumbered the negative ones and six learners were not able to recall any negative assessment-related situations, it can be concluded

that, on the whole, the participants held rather positive views of language assessment they had experienced during their education and perceived the classroom assessment environment as safe and comfortable. Both the reports of critical incidents and the accompanying justifications suggest that the learners trusted their teachers, believed in their professional competence and felt confident that their efforts were adequately appreciated. The fact that only one critical incident referred to an assessment practice in which the principle of validity was violated indicates that the learners viewed their assessment as congruent with the learning objectives and the content of teaching. These findings stand in contrast to the research conducted on a group of Chinese high school learners, in which negative perceptions of assessment, stemming from perceived heavy workload and unsatisfactory results, prevailed (Xiao & Carless 2013).

It is clear that reliability is believed to be the most important aspect of assessment. Learners require that assessment results fairly reflect the effort they make and any cases of violating this principle were instantly interpreted as sources of injustice and teacher's bias. The second most frequently recalled aspect was appreciation, which is unsurprising, given the participants' stage of social and emotional development. Adolescence is a period of shaping one's self-concept, which depends on individuals' interaction with the environment and the effect they have on others. Due to the growing self-awareness and the willingness to establish their position in a group of peers, adolescents are very sensitive about the way they are perceived by the surrounding world and, therefore, their self-esteem depends on social acceptance and positive reinforcement from significant others, including teachers (Coleman & Hendry 1990).

Very few references pertaining to the practicality category imply that the learners seemed to be ready to work hard and complete the assigned tasks, provided their workload was distributed reasonably and their efforts were properly appreciated. It must be noted here that even though task-based tasks are normally regarded as time- and work-consuming, some respondents enjoyed them as they were believed to be as meaningful, authentic and involved some degree of independence.

The respondents' view of assessment may be largely positive; still, it is worrying that the respondents paid relatively little attention to the learning gains that could be derived from this aspect of language teaching. Whereas a few learners underlined the importance of meaningfulness, authenticity and formative potential of assessment, a vast majority tended to put more weight on receiving a good grade, being appreciated by the teacher and comparing their own results with those obtained by other learners. Such

overemphasis on grading and praising, often at the cost of self-improvement and formative feedback, is characteristic of traditional and grade-oriented systems of education. Looking back at Alkharusi's (2011) two-factor scale recalled in the theoretical part of this article, it can be concluded that performance-oriented assessment environment dominates in the investigated context. There seems to be a symbiotic relationship between how learners perceive assessment environment and their teacher's approach to giving corrective feedback. It is evident that in many cases the teachers used holistic assessment, and the learners' overall performance was summarized with a single grade, sometimes accompanied by a general comment. Moreover, on a few occasions, the learners were not informed about assessment criteria either before or after a particular task was assigned, which makes it hardly possible for the learners to actively participate in the assessment process and fully benefit from the feedback they receive. As McNamara (2000) underscores, assessment can be regarded as ethical only when test-takers are provided with a detailed description of elicitation tasks, content, criteria and standards. Undoubtedly, the presentation and explanation of the objectives and assessment criteria of each assigned task would support learning, eliminate unnecessary misconceptions concerning language assessment and increase learners' awareness of the benefits of formative feedback.

It is additionally worth analysing whether and to what extent CIT appeared to be useful in eliciting adolescent learner perceptions of language assessment. The first advantage of this technique was that the participants were afforded freedom of choice: from a wide repertoire of assessment experiences, they could choose the ones they subjectively considered as particularly memorable and important, both in a positive and negative sense. Such an approach to eliciting data helped to analyse learner perceptions of assessment from a more holistic perspective. Whereas earlier studies explored this issue in reference to one pre-selected assessment mode, for instance a high-stakes test (Huhta et al. 2006) or portfolio assessment (Czura 2013), the application of CIT in the present study created the possibility of exploring which aspects of assessment learners perceived as particularly important and what, in their opinion, makes a good assessment. On the negative side, the idea of a critical incident was not immediately understood by some learners, which required the researcher to provide instruction that would be more detailed and informative, and, at the same time, would not direct learners' attention to a particular assessment technique.

The main limitation of this study is its small sample size. Since the study involved a group of 25 adolescents attending the same school, the results cannot be generalised to other populations and settings. On the other hand,

it must be noted that it is the immediate learning environment, and not the educational context at large, that has a dominant impact on learner perceptions of learning and studying (Entwistle 1991). This phenomenon is additionally supported by the results of a large-scale study conducted by Cheng, Wu and Liu (2015), which revealed that university student perceptions of assessment differed depending on the university, the study programme and self-perceived level of language proficiency. In future studies on learner perceptions, it might be interesting to compare the findings of the present study with the results of investigations carried out in other schools and educational contexts. Future researchers could also design studies to find correlations between the learners' perceptions of assessment and their final attainment, level of motivation and self-efficacy.

## 6. CONCLUSIONS

The objective of this study stemmed from the dearth of empirical data on learner perceptions of foreign language assessment. The results reveal that the participants held rather positive views of language assessment they had experienced during their education and paid most attention to such aspects of assessment as reliability, appreciation and authenticity. On the other hand, it is evident that the importance of grades was prioritised over the learning value of assessment, which is characteristic of performance-oriented assessment environment (cf. Alkharusi 2011). Even though, due to a small sample size, the results cannot be generalised to other contexts, the findings of the study may be treated by foreign language teachers as indicators of what learners in this age range regard as positive or problematic in language assessment. The analysis of learners' accounts of both positive and negative critical incidents may prompt the teachers' and teacher educators' reflection on day-to-day assessment practices and approaches to corrective feedback. It seems necessary to both cultivate the examples of good practice that already exist in Polish schools and find ways of maximising the formative potential of classroom assessment.

## REFERENCES

- Alkharusi, H. (2011). Development and datametric properties of a scale measuring students' perceptions of the classroom assessment environment. *International Journal of Instruction*, 4(1), pp. 105–120.

- Alkharusi, H. (2015). Classroom assessment communication, perceived assessment environment, and academic achievement: A path analysis. *British Journal of Education, Society & Behavioural Science*, 8(2), pp. 117–126. DOI:10.9734/bjesbs/2015/17670.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Carless, D. / Lam, R. (2012). The examined life: perspectives of lower primary school students in Hong Kong. *Education*, 42(3), pp. 313–329. DOI:10.1080/03004279.2012.689988.
- Chell, E. (1998). Critical incident technique. In: G. Symon / C. Cassell (ed.), *Qualitative methods and analysis in organizational research* (pp. 51–72). London: Sage.
- Cheng, L. / Wu, Y. / Liu, X. (2015). Chinese university students' perceptions of assessment tasks and classroom assessment environment. *Language Testing in Asia*, 5(1), pp. 1–17. DOI:10.1186/s40468-015-0020-6.
- Coleman, J. / Hendry, L. (1990). *The nature of adolescence*. Second edition. London: Routledge.
- Czura, A. (2013). Implementing portfolio assessment in lower-secondary school. *ELOPE: English Language Overseas Perspectives and Enquiries*, 10(1), pp. 83–94. DOI:10.4312/elope.10.1.83-94.
- Dorman, J. P. / Knightley, W. M. (2006). Initial use of the Perceptions of Assessment Tasks Inventory (PATI) in English secondary schools. *Alberta Journal of Educational Research*, 52(3), pp. 196–199.
- Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. Introduction to the special issue. *Higher Education*, 22, pp. 201–204.
- Farrell, T. S. C. (2008). Critical incidents in ELT initial teacher training. *ELT Journal*, 62(1), pp. 3–10.
- Finch, A. (2010). Critical incidents and language learning: Sensitivity to initial conditions. *System*, 38(3), pp. 422–431.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), pp. 327–358.
- Gabryś-Barker, D. (2012). *Reflectivity in pre-service teacher education. A survey of theory and practice*. Katowice: University of Silesia Press.
- Huhta, A. / Kalaja, P. / Pitkänen-Huhta, A. (2006). Discursive construction of a high-stakes test: the many faces of a test-taker. *Language Testing*, 23(3), pp. 326–350.
- Kuchuk, A. (2012). *Politeness in intercultural communication: Some insights into the pragmatics of English as an international language* (Unpublished doctoral dissertation). University of Arizona, Tucson.
- Lizzio, A. / Wilson, K. (2013). First-year students' appraisal of assessment tasks: implications for efficacy, engagement and performance. *Assessment & Evaluation in Higher Education*, 38(4), pp. 389–406.
- McAllister, L. / Whiteford, G. / Hill, B. / Thomas, N. / Fitzgerald, M. (2006). Reflection in intercultural learning: Examining the international experience through a critical incident approach. *Reflective Practice*, 7(3), pp. 367–381.
- McClure, J. W. (2007). International graduates' cross-cultural adjustment: experiences, coping strategies, and suggested programmatic responses. *Teaching in Higher Education*, 12(2), pp. 199–217. DOI:10.1080/13562510701191976.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Messick, S. (1989). Validity. In: R. Linn (ed.), *Educational measurement* (pp. 13–103). New York: American Council on Education.
- Pollard, A. / Triggs, P. (2000). *What pupils say: Changing policy and practice in primary education*. London & New York: Continuum.

- Stiggins, R. J. / Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.
- Struyven, K. / Dochy, F. / Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review1. *Assessment & Evaluation in Higher Education*, 30(4), pp. 325-341. DOI:10.1080/02602930500099102.
- Tripp, D. (1993). *Critical incidents in teaching: Developing professional judgment*. New York: Routledge.
- Xiao, Y. / Carless D. R. (2013). Illustrating students' perceptions of English language assessment: Voices from China. *RELC Journal*, 44(3), pp. 319-340.

**Received:** 26.01.2017; **revised:** 17.03.2017

