Professional Quality Assessment of the Croatian State Written Exam in Biology

Ines Radanović1, Žaklin Lukša2, Valerija Begić3, Marijana Bastić3, Diana Garašić1, Mirela Sertić Perić1, Ivanka Podrug3

1Faculty of Science, Zagreb, Croatia
2High School Josipa Slavenskog, Čakovec, Croatia
3Primary school, Zagreb & Split, Croatia

The aim of the tool presented in this study is to enable teachers’ qualitative analysis of the questions within the Croatian state written exam in biology, and the eventual correction of the questions before their application in the student assessment. We have identified the two basic categories that determine the question quality: 1) the importance of questions (regarding the profession, life, curriculum, critical thinking), and 2) the influence of questions on students’ answers (i.e., type, shape and intelligibility of the questions as well as logical reasoning and further learning path in relation to the students’ answers). The tool we have developed was tested for its effectiveness on a sample exam designed for students aged 13. A correlation between logical reasoning and the ‘importance-of-questions’ categories, and the success rate of the exam was observed. This simple tool has proven to be effective for both teachers’ self-assessment and peer evaluation.

Keywords: cognitive levels, question relevance for science literacy, influence of questions to answers

Introduction

Most classroom teachers prepare and administer a series of formal (i.e., teacher-made) exams during the school year, which often enclose questions with many construction mistakes, especially essay questions (e.g., Marso & Pigge 1988). Thus, there is a growing need for greater quality control in the design and execution of the students’ performance assessments (Dunbar et al. 2009). A tool for the expert question quality assessment in Croatia (i.e., a country still developing its science literacy and national curriculum frontiers) was for the first time designed for the needs of professional quality assessment of state exams in biology (Radanović et al. 2010). In designing the Croatian tool, the following criteria, recognized as “fruitful areas” to seek question validity evidence, were considered: question content, internal structure and response process, as well as exam scores’ relationship to other variables measuring various students’ domains, and overall learning success and achievement (Downing, 2003). From its first use, the Croatian tool has been continuously developed through the application within research, as well as within teaching, i.e., in designing written biology exams (Radanović et al. 2011, Begić et al. 2016). As for the purpose of the tool development, we introduced some assessment elements that should encourage teachers to better prepare exam questions, the question quality has steadily increased since the launch of the tool. The aim of the tool we present here is to enable teachers’ qualitative analysis of the questions within the Croatian state written exam in biology, and the eventual correction of the questions before their application in the student assessment. An additional aim is to enable the qualitative question analysis in order to more comprehensively interpret student results within the written exam.

Method

Based on years of experience in the usage of the question analysis with the assistance of experienced biology teachers, we have developed the tool for assessing the quality of biology state written exam questions. Elements and criteria for the expert question quality assessment (Table 1) were determined by three point Likert Scale (Cohen et al. 2007). By harmonizing the statements using an unambiguous numerical scale (Table 1), it was possible to make a comprehensive question quality assessment, which was
initially unfeasible because of the two adverse scales assessing the importance of questions, and the influence of questions on students’ answers separately by averaging the scales’ scores (Table 1).

Table 1. Elements and criteria for the expert question quality assessment.

<table>
<thead>
<tr>
<th>Question quality</th>
<th>The importance of questions</th>
<th>The influence of questions on students’ answers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Assessment elements of the science literacy</td>
<td>Scale of the question importance</td>
</tr>
<tr>
<td>1 – BAD</td>
<td>A – importance of questions for the profession</td>
<td>1 – unimportant</td>
</tr>
<tr>
<td>2 – ACCEPTABLE</td>
<td>B - importance of questions for life</td>
<td>2 – moderately important</td>
</tr>
<tr>
<td>3 – GOOD</td>
<td>C - importance of questions for the curriculum</td>
<td>3 – important</td>
</tr>
<tr>
<td></td>
<td>D – importance of questions for critical thinking</td>
<td>(Qim+Qin)/2</td>
</tr>
</tbody>
</table>

By the assessment of the importance of questions, the following items are scored: A – important for the development of basic biological concepts (conceptual development); B – important for basic biological literacy and life applicability (as context-rich questions); C – important for the development of the competences foreseen by the curriculum; D – important for the design and implementation of natural science methods or analysis/synthesis during the initial information sorting. By the assessment of the influence of questions on students’ answers, the following issues are taken into account: A – technical characteristics of the question (information necessary to solve the task); B – adjustment of the question to students (without distractors, e.g., too many technical terms, etc.); C – possibility of answering the question without understanding the overall concept; D – does the question require further learning, as it examines knowledge beyond the prescribed curriculum or the details irrelevant for conceptual understanding. The effectiveness of the developed tool was estimated by 4 teachers by means of 23 biology exam questions (Cronbach’s alpha = 0.583, n = 148, SE = 0.048, 95% CI = 0.48 to 0.67) targeted for students aged 13 (Begic et al., 2016). Statistical analysis was done by StatsToDo (Chang, 2014), and correlations are interpreted according to Hopkins (2000).

RESULTS

Spearman Rank Order Correlation was proven significant for logical reasoning and the “importance-of-questions” categories in relation to the success rate of the exam (ρ = 0.44, p < 0.05). Concordance as a measure of agreement between opinions indicated a weaker concordance of the reasoning among students (average Fleiss kappa = 0.32). There was a greater concordance among teachers regarding the assessment of the importance of questions (Kendall W = 0.53; ChiSq = 46.42; df = 22; p = 0.001) than regarding the assessment of the influence of question on students’ answer (Kendall W = 0.31; ChiSq = 27.11; df = 22; p = 0.21). According to the final question quality assessment done by averaging the scales’ scores (Table 1), there were no statistically significant differences between the individual teachers’ assessments (Kruskall-Wallis H = 0.25; df = 3; p = 0.97), and the teachers were relatively well-matched in their assessments (Kendall W = 0.44; ChiSq = 39.08; df = 22; p = 0.01). Authors of the questions were shown to be less self-critical in the self-evaluation than their peers, but this difference in the self-assessment was very low (8.6%).

DISCUSSION AND CONCLUSIONS

Already during the initial application of our tool, it was noted that the critical assessment of the elements and criteria (Table 1) coincide with the results of psychometric question analysis (Radanović et al., 2010). Quality of the questions has lasting effects on teaching and learning, so the technical properties of the questions should be greatly considered by developers and practitioners (Dunbar et al. 2009). Discordance
among the teachers' assessments confirms that the teachers are not prone to critically reflect on the questions they shape. The post-hoc quality analyses of their exams, based on the consensus (i.e., averaging) among the selected elements and criteria for the question quality assessment, would provide plenty useful and relevant information on the overall question quality (Marso & Pigge 1988). More uniform teachers' assessments of the importance of questions might confirm the teachers’ competence within the subject (i.e., biology), their knowledge and professional expertise. We suggest that the final assessment of the effectiveness of written exam (i.e., questions) should be done “by science” (Begić et al., 2016), whose members are the most relevant for assessing the higher cognitive levels through, for example, the open-ended questions, which have high content validity, but rather poor inter-rater objectivity (Schmelzing et al. 2013). The teachers’ disagreement in the assessment of the influence of questions on students’ answers could indicate an uneven teaching experience. Such result suggests that the teachers should necessarily continuously work on their own professional development (Gottheiner & Siegel 2012) to be able to focus well on setting the question quality standards (e.g., technical preparation of the questions, adaptation of the questions to the students, avoiding questions that demand high level of logical thinking, etc.). The teacher professional development should further help teachers to close the formative assessment cycle by addressing conceptions that are elicited with assessments (Gottheiner & Siegel 2012). Additionally, there is a need to develop the result analysis criteria for the exams, and a scientifically based approach to their assessment (Golovachyova et al. 2016).

The tool we developed could be used for peer-evaluation as well as for self-assessment, but only if critically applied with the recommended delay of at least 2 weeks after the question preparation. The most important roles in the question quality assessment play the teachers’ experience in the classroom as well as the overall experience in the question analysis. Therefore, it is very important to encourage teachers to apply collaborative quality assessment approach.

REFERENCES


