

# Testing & Evaluation & Assessment

## TEASIG Webinar Series

### Language assessment is NOT a solitary activity: The benefits of teachers and testers' collaboration for all assessment stakeholders

Anna Soltyska 09/6/20

*As time did not allow for all the participants' questions to be answered during the webinar, Anna has provided comments and answers below.*

Thank you for inspiring and reflective questions submitted during the webinar. As I already addressed several of them in the Q and A session after my talk, I have tried to combine similar questions (or those aiming at similar concepts) and answered them collectively.

**What makes a successful collaboration? What are the parts of collaboration that prove a lot more challenging than others? What about a neutral party within the collaboration team to ensure validity and mutual respect? What are the challenges of collaborating online?**

Collaborating online is not much different from face-to-face collaboration, though some disparities might affect one's preference for or aversion to this particular mode of cooperation.

While searching for potential partners to collaborate conventionally, i.e. including meetings to discuss matters in person or working together on paper documents, an argument of geographical proximity or convenient travel connections might limit the choice of collaborators. However, there seem to be few constraints of this kind if meetings and communication are to take place online. Nevertheless, the more global and heterogeneous a membership of a project, the more relevant issues of different time zones and accessibility of collaborative tools in various countries might become. Access to technology including software (some programs require costly licenses, others might entail complicated registration procedures) and hardware is a further objective factor that might pose a limitation to collaborating online and certainly affect the choice of collaboration mode (emails? instant chats? video conferences?...), which in turn might impact collaborative processes and ultimately the desired outcomes. A further consideration is the security of data (e.g. the need to anonymise all student work) when processing it online and relying on cloud services. Here local regulations might differ from country to country (or institution to institution) and severely restrict collaborative work.

As far as human factors are concerned, as in any collaborative project (also in those carried out online) divergent styles of working and interacting within the team – among other reasons due to various cultural and language backgrounds, but also occupational and family status – are likely to challenge the endeavour. Establishing clear objectives, identifying tasks to be performed and reaching a consensus with regard to what is expected of everyone in terms of availability, responsibility and timely deliveries is necessary for any collaboration to be successful. If needed and accepted by all participating parties, a neutral ‘third party’ with a supervisory role might be invited to oversee progress, ensure validity and report to further stakeholders (e.g. sponsors or patrons of the project). The online working mode might magnify the shortcomings in the project management, for instance if overall progress is hindered by individuals struggling with technology or failing to engage themselves in online discussions, though such projects might well flourish when well-scheduled and under skilful and efficient management.

It should also not be forgotten that – in its most orthodox variety, i.e. ‘online only’ – the reduced personal contact with prospective and current collaborators might prove irreplaceable and result in unpredicted consequences for the projects and their stakeholders.

**Have any changes been introduced (in both universities participating in this collaboration) based on the project findings? Are the results (of the project) applicable worldwide?**

The results of the calibration project, which was carried out in two German university language centres over 12 months and focused on the assessment of the skill of writing at level B2 according to CEFR, are definitely worth being analysed and applied by language teachers and testers active in other contexts. The initial findings of the first phase (analysis of 67 essay topics) led to identifying essential features of a typical B2 essay writing task. Characteristics of the task such as a context, setting, target group, genre, etc. were defined and several skills (German ‘*Handlung*’) that a task might involve were listed. These findings were then implemented when drafting essay topics for the exams and revising the assessment rubrics. Secondary findings (raters’ reflections on assessment criteria applied to grade students writing samples – 32 texts in total were analysed in both language centres and assessed independently by at least two raters based in the same institution) served as basis for optimising the assessment rubric and standardising the pass-threshold.

Both institutions have clearly benefitted from the project in terms of increased assessment literacy within their teams, and standardization of their assessment procedures by ‘sneaking-a-peek’ into the assessment practices of another institution. Furthermore, as concerns of comparability, inter-rater reliability and rater bias were discussed and closely analysed in the course of the project, all stakeholders (including the decision-makers in both institutions) reaffirmed their interest in maintaining the high quality of assessment, including through the procedure of double rating, at whatever cost. These are the lessons that are relevant and valid not only for the participating teachers, testers, raters and their institutions, but for the language testing community as a whole.

*(An article that will present the project and its outcomes and implications for those involved as well as other middle-stake contexts is being prepared for publication.)*