IO5 REPORT | UV

OUTPUT IDENTIFICATION: O5	
OUTPUT TITLE: Evaluation and Testing	
START DATE:	01-11-2016
END DATE:	31-08-2019
LANGUAGES:	English
MEDIA(s):	- Text - Publications - Handbook - Oral - Event - Other

OUTPUT DESCRIPTION:

Testing and checking for quality and striving for excellence is one of the driving forces in the ADLAB PRO project. The project intends to plan a strategic development of course content, derived from IO1 and IO2, but also checking the efficacy of the course content and design with actual potential users during IO4. These evaluations are labelled for this project as "testing" but in fact the IO goes beyond this. Each course content from IO4 has been designed in a progressive form - rather than in parallel. This has allowed for testing the content at each stage, and also to guarantee that the progress and evaluation are correct. Any feedback has been implemented immediately - hence avoiding a final correction of all material. This sequential development with ad hoc testing has secured content quality, adequacy and progress. The Partners recognize the importance of evaluation such that it has been built in from the very beginning of the project. This is why evaluation has been listed as one of the intellectual outputs and thought about from the start. The reason to start with the project onset, was to develop test content, methodology and Key Performance Indicators (KPIs) or Quality Indicators (QIs) for tests to be ready in time for implementation.

In addition to user satisfaction, Sweller et al. (1978; 1988) note the importance of testing cognitive load (CL) for training materials, such that the materials produced are neither too demanding, nor insufficiently demanding, using standardized z scores to overcome differences in individuals' performance. ADLAB PRO has also tested CL, via a simple subjective measure of mental effort, using a 9-point Likert scale (where 1 = very, very low mental effort and 9 = very, very high mental effort). This has been shown by Paas (1992) to be an instrument which is surprisingly sensitive and to correlate strongly with more complex physiological interventions such as spectral analysis of heart rate. In this way the different training resources produced as part of IO4 may be reliably ranked according to cognitive demands, such that trainers can select material appropriately for learners in different settings (universities, industry etc.). This also means that resources can be assessed as to where they fit within the standards and frameworks created by IO6.

The partners recognise the importance of incorporating the feedback between workshops to refine the materials, improving them if necessary before the next one. This means that longitudinal data will also be available, tracking the evolution of each individual resource that the Project will produce. The evolution of each material type is tracked in this report.

In addition to quantitative testing, a strong emphasis has been placed on qualitative measures such as written and recorded verbal feedback arising from semi-structured interviews to check that participants found the resources engaging, stimulating and helpful in advancing their knowledge of accessibility in general and AD in particular. It was envisaged that testing would also use quantitative feedback including the use of measures currently used for reception studies in AD and subtitling, such as the ITC SOPI (Lessiter et al., 2001) which is focused on user experience (UX) (Fryer & Freeman, 2014). In the event UX measures were selected that assessed learning and reaction stemming from the model proposed by Kirkpatrick (1959) specifically for the evaluation of training materials. These measures included interest, attention, confusion and ease of understanding as well as satisfaction with the way contents were structured and presented. These seemed more appropriate than the ITC SOPI which measures immersion and was designed for assessing UX in the sphere of entertainment such as online gaming.

INTRODUCTION

The aim of IO5 was to test and evaluate the content developed by ADLAB PRO at each stage of the project to guarantee smooth and effective progress. It ensured that any feedback would be implemented immediately to avoid a final correction of all material. This sequential development combined internal and external evaluation, to test the efficacy of the course content and design with actual potential users during and after IO4.

The external evaluation central to IO5 has taken the form of a series of short reports, combined into a handbook called A Guide to the Evaluation of Training Materials: ADLAB PRO: A Case Study (henceforth The Guide) which is available here: [link to be added once the Guide is uploaded]. The Guide is a compendium of the evaluation that has underpinned the development of the training materials produced under IO4, showing their evolution from earlier IOs, as well as a comprehensive discussion of the theory on which it was based. It is envisaged that The Guide will provide a useful resource for AD professionals as a summary of available training with an indication of its reception by different audiences who are all stakeholders in AD. Beyond this, The Guide will be useful for any researcher considering evaluation with a thorough discussion of evaluation strategies, their advantages and limitations. Some of its content is reproduced in this final report. However, readers interested in the detailed analysis of the data acquired for IO5 are referred to The Guide. It was envisaged in the application that The Guide would double as the final report. However, owing to its length and complexity, a digest is presented here for ease of reading.

This report concentrates on the training materials produced for IO4. It explains how they have been assessed, demonstrating their evolution and how they have been refined throughout the project. An overview of the methodological approach is followed by a summary of results following the various evaluations of each material type (Core videos; Reading Lists; Additional Videos; Trainer's Guides; Tasks and the materials in general). It is followed by an analysis of the strengths and weaknesses of the materials and of the evaluation process. The result is a summary of the qualitative and quantitative data that has been obtained from satisfaction surveys and measurable evaluation indexes created for ADLAB PRO.

WHAT IS EVALUATION?

Wigley (1988, p.21) defines evaluation as "a data reduction process that involves the collection of large amounts of data which are analysed and synthesized into an overall judgement of worth or merit." Marsden (1991, p.31) found that "Evaluation is given a low priority in the instructional process, a contention that is supported by the small number of articles in the literature that deal with it." She argues that evaluation should be given a high priority as it provides evidence that justifies the value and viability of training programmes. "Yet evaluation is often something of an afterthought for those whose main concern is with delivering training" (1991, p.31). Eleanor Chemlinsky (1997) distinguishes between three broad types of evaluation: for accountability (to funders and other stakeholders); for causal knowledge "to generate strong evidence that the intervention causes the intended outcomes" and for program improvement. These broadly fall into one or other of two types of evaluation: formative and summative, each with a distinct aim.

FORMATIVE EVALUATION

The aim of formative evaluation is to improve an outcome as it develops. For example tasting a dish while it is cooking in order to decide whether or not to add more seasoning. Formative evaluation was at the heart of the evaluation process for the IO4 materials.

SUMMATIVE EVALUATION

The aim of summative evaluation is to assess the final result. For example tasting a dish once it is ready to determine how well it has been cooked and to decide whether or not to use that recipe again. Summative evaluation allows the project to check that it is meeting its goals; to better communicate its achievements; to acknowledge any serendipitous gains or unexpected insights; to show what the project has done and to celebrate what it has achieved. This report can be thought of as a summative evaluation of the Intellectual Outputs (IOs) of ADLAB PRO.

TIMING

- The evaluation cycle

Scheirer and Schwandt (2012) emphasise that evaluation is necessary at every stage, while recognising that in reality these stages are not always discrete. Frechtling (2002) proposes a project planning and evaluation cycle such that after the project is planned, an evaluation process establishes the current situation, the status quo. This type of evaluation will collect baseline data from specialists and perhaps conduct a literature review at which point the project solution is modified if necessary before being implemented in a pilot phase. Once implemented, the project is reevaluated, by comparing post-project implementation data with the baseline, allowing for any differences to be measured. In this way, the effect of the project can be ascertained. Summative evaluation is a one-off event at the final stage of the project. However, formative evaluation can take place repeatedly throughout the project's life to ensure component parts are satisfactory and in order to avoid any nasty shocks at the end. As regards ADLAB PRO: Before building the AD curriculum, and long before designing any course materials, the project began by assessing current AD training practices (IO1). Once this IO was completed, it was itself evaluated as part of the project's internal evaluation process (see section 2.2.1), before moving on to the next IO (IO2).

- Planning

Evaluation must be thought about in advance with sufficient time given not only for the evaluation to take place but also for any findings or modifications to be implemented before that particular phase of the project is deemed complete and the project is allowed to move on. It should be noted that any extensions to project completion dates also have budgetary implications. Frechtling (2002, p.46) suggests 5 – 10% of the overall budget as a ballpark figure for evaluation costs. Time should be allowed for identifying potential evaluators; contacting potential evaluators; substituting evaluators in case of illness or unwillingness to take part; giving evaluators sufficient time to complete the evaluation; collecting and analysing the completed evaluations; writing an evaluation report and implementing changes where necessary. This also raises the question as to who evaluates? This is addressed in the next section.

WHO EVALUATES?

Wilkes & Bligh (1999) suggest that evaluations in education can take one of four orientations. They can be oriented towards stakeholders but they might equally be oriented towards students; the course or the

institution. Each requires a different indicator such that student oriented evaluations rely mostly on measures of student performance whereas a programme oriented approach compares how the course performs with respect to its stated objectives and often involves descriptions of curriculum or teaching activities. It evaluates how elements of the course have contributed to student outcomes. A Stakeholder oriented approach takes into account the concerns and claims of those involved in and affected by the course (stakeholders in ADLAB PRO are identified and discussed in section 1.4.1 below).

For ADLAB PRO, an institution-oriented approach was deemed inappropriate, as the consortium has control over neither the quality of teaching nor the institution where the training takes place. ADLAB PRO has control only over the quality of the ADLAB PRO materials. A literature review pertaining to Quality in AVT resulted in a publication for IO5 (Fryer, 2019). A programme-oriented approach was also deemed to be problematic, as the materials have been designed to be flexible and modular such that they could either form a complete course, or offer a "pick'n'mix" selection from which a trainer could supplement their own materials. Any complete course evaluation would need to take account of the role played by materials external to the project.

In many disciplines, the ideal form of evaluation is thought to be a randomised control trial (RCT) of the type much prized in medical studies. Schreier and Schwandt note that "the extensive set of criteria for a valid RCT is not likely to be present within small non-profit agencies that are often the recipients of human service program grants" (2012, p.8) and conclude that a generalisable causal model is not feasible in such cases. In addition, a programme-oriented approach is necessarily a form of summative evaluation. Students attending a course could not ethically be taught exclusively using materials that had not been tested, nor could a complete testing of all the course materials take place until the very end of the project. For these reasons Utopian Voices (UV), decided to combine stakeholder, student and programme oriented approaches, but with the emphasis on a stakeholder approach for the formative evaluations. By combining approaches, evaluations can be triangulated from different perspectives, adding to the reliability of the results.

- Stakeholders in ADLAB PRO

According to Mitchell, Agle and Wood (1997, p.854) "stakeholder" is a term stemming from management theory. It was popularised by Freeman (1984) who defined stakeholders broadly as "any group or individual who can affect or is affected by the achievement of the organization's objectives" (1984, p.46). Mitchell et al. (1997) also cite Windsor and his narrower definition of stakeholder groups as "those on which the organization is dependent for its survival" (1992, p.19). Clarkson narrows this still further to a division between "voluntary and involuntary risk-bearers." For example, in ADLAB PRO, the EU as the funding body has taken a voluntary risk by investing capital in the project, while people with sight loss (PSL) are involuntarily "placed at risk by the project's activities" (Clarkson, 1994, p.5) as they might suffer from exposure to poor description if describers trained using ADLAB PRO materials fall short. If that sounds alarmist, Clarkson asserts "without the element of risk there is no stake" (1994, p.5). Further categories of people with a stake to lose or gain from ADLAB PRO's activities include students wanting to learn the skills of audio description and anyone concerned with course delivery, that is teachers and trainers in a variety of settings and by extension, as stated above, anyone concerned with the AD produced by graduates of the course, namely AD providers and PSL, who are principally the intended beneficiaries of AD, and their advocates including organisations concerned with promoting equality of access. The ADLAB PRO training

materials could be likened to a stone thrown in a pond, rippling out to various stakeholders as shown in Fig. 2.

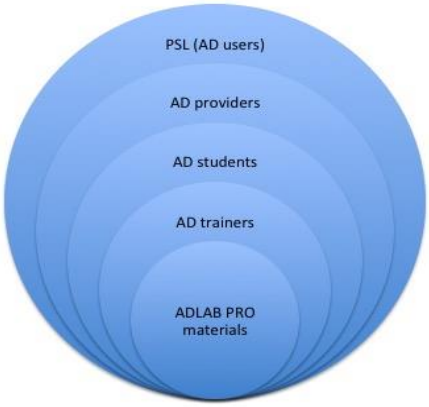

Figure 2. Stakeholders in ADLAB PRO

After a Partner discussion at TPM 2, it was decided that the primary stakeholders are AD trainers, a term which includes lecturers and teachers, in short anyone in any learning environment wanting to deliver a course in AD. The secondary stakeholders are the students who learn AD skills from these trainers, as the materials are not primarily designed for self-learning. Once students have completed their training and acquired the competences identified in IO1, confirmed in IO2, and specified in the ADLAB PRO curriculum (IO3) they will be able to make their skills available to AD providers. Providers might include a TV company wanting to broadcast its output with AD, a theatre company or venue wanting to fulfill its obligations under the European Accessibility Act (2018) or an agency or facilities house that provides access services to a variety of clients. Alternatively a newly qualified describer might set up their own company to provide AD services or volunteer for a charity or an NGO engaged in providing AD. Whichever way, the ultimate stakeholders are the PSL who will make use of the AD provided to engage with cultural products including (but not limited to) film and TV, live performances, heritage buildings and artworks.

- Influencers

Mitchell et al. (1997, p. 859) further distinguish between claimants and influencers whereby influencers "have power" whether or not they have any claims at all. In the case of ADLAB PRO, PSL might be deemed to have the strongest moral claim on the output of the project, with students and trainers having a little less, but all three might be thought to differ in the amount of influence they bear. For example, trainers might be thought to bear the greatest influence. As their opinions of the materials will determine whether or not they use them, or recommend them to others, trainers will have a major influence on the success and sustainability of the project. Trainers, in turn, will likely be influenced by student opinion, for they are unlikely to continue to use materials if students find them too complex, too simple or simply unengaging.

- Evaluators for ADLAB PRO

Partners settled on three groups of stakeholders to act as evaluators at various stages in the development of the IO4 materials: AD trainers, students and PSL.

Before moving on, it should be noted that just as the stages in the evaluation cycle are not discrete, nor are the roles of stakeholders. IO1 sampled 86 AD teachers and trainers and found that 55% of them also identified as audio describers (ADLAB PRO, 2017a). Similarly some trainers and a few audio describers are blind or partially sighted (Fryer & Cavallo, 2018).

HOW? EVALUATION MEASURES

Having determined who should evaluate, the next important decision was to determine how the evaluations should be carried out. The main division in measures of evaluation is drawn between qualitative and quantitative methods. These are fully explored in The Guide. As both have limitations it was decided to adopt a mixed measures approach.

- Mixed measures

In contrast to the disputed pros and cons of the two broad categories of evaluation methods, a third way combines quantitative and qualitative. This "mixed methods" approach enjoys both the depth of quantitative methods and the breadth of qualitative methods and has been gaining ground in numerous disciplines including education, psychology, social and health sciences (O'Cathain, 2009). O'Cathain argues that this increase reflects the complexity of health interventions, such that qualitative methods enable quantitative results to be interpreted. She cites a report by the UK Medical Research Council (2000) that advocates the use of qualitative methods in the early stages of an intervention to understand the subject of study, and again at the end to help understand why an intervention did or did not work. This would be equally applicable to the needs of educational research. O'Cathain also cites examples of what are called intermethod discrepancies. For example Campbell, Quilty & Dieppe (2003) report instances where people who talked about an improvement in their health did not show an improvement on the quantitative scale. Amongst AD learners it may be imagined that after refresher training in AD, a describer may feel more confident in their abilities, although their AD may not be assessed more highly. One reason for the use of multiple methodologies is to control systematic biases. However, Collins, Onwuegbuzie & Jiao (2006) point out that reasons for under-use of a mixed methods approach include the greater labour intensity. They conclude that combining approaches requires "more time, resources and effort to organize" (2006, p.68) as well as greater expertise to design and implement such studies.

ADLAB PRO adopted a mixed methods approach with the balance tipped towards qualitative data in the formative stages. This was because the number of responses was likely to be small and suggestions for improvement were more helpful at the early stage of the project than obtaining statistically significant results. In particular Partners were keen to absorb suggestions from key informants on ways to improve the materials. Partners were also aware that triangulation from different methods leads to thicker, richer data allowing stakeholders to be more confident of the results.

QUALITY INDICATORS RELEVANT TO TRAINING MATERIALS

60 years ago, Donald Kirkpatrick (1959) proposed materials for management training programmes be evaluated on the basis of reaction; learning; behaviour and result. Mahapatra and Lai (2005, p.68) expand further on Kirkpatrick's ideas to explain that reaction "measures the satisfaction of the trainee with the learning material"; learning "measures the skill or knowledge learned"; behaviour refers to the "effect of the learning on the trainee's job performance" while result is a global outcome namely the "effect of the training programme on overall organizational performance". While Kirkpatrick's model takes an institution-oriented approach that does not entirely fit with the organisational set up of ADLAB PRO nonetheless the first two of his broad principles were a good fit and have formed the basis for the indicators chosen.

- Reaction

Reaction was evaluated using self-perceived measures of interest, attention, confusion and ease of understanding as well as satisfaction with the way contents were structured and presented. Qualitative comments solicited what the respondent like best and least about the learning materials. One global measure of satisfaction was to ask whether or not respondents would recommend the training materials.

- Learning

Learning was also self-assessed. Participants were asked the degree to which they felt the materials had developed their skills or increased their understanding of AD. At the formative stage trainers were asked to assess what they thought the response of their learners would be. Later, once all the materials were complete, a multiple choice task that was integral to the IO4 Tasks provided a direct measure of comprehension. In addition learners were asked to rate the amount of mental effort they expended in following the materials. This was to give some idea of the cognitive load (CL) a concept relating learning to cognition that was developed some decades after Kirkpatrick came up with his model.

- Cognitive Load

CL as defined by Sweller (1988) can broadly be defined as the amount of mental effort it takes to process information. It has particular application to learning materials as it integrates components of cognitive architecture which might affect how an individual apprehends information. For example tasks may overload the learner if the load it places on their working memory (WM) exceeds the individual's capacity. This will depend on a combination of the individual student's WM, circumstances beyond the task itself such as classroom conditions (extraneous load) and demands that are intrinsic to the task. De Jong (2010, p.106) explains that "intrinsic cognitive load relates to the difficulty of the subject matter. More specifically, material that contains a large number of interactive elements is regarded as more difficult than material with a smaller number of elements and/or with a low interactivity. Low interactivity material consists of single, simple, elements that can be learned in isolation, whereas in high interactivity material individual elements can only be well understood in relation to other elements (Sweller 1994; Sweller et al. 1998)."

Creators of learning materials must ensure that task demands are neither too high (overwhelming WM) nor too low (which leads to boredom and loss of attention). The ideal amount of CL is termed germane load, although it is hard to quantify the precise value as it is an interaction between extrinsic and intrinsic aspects of the materials and the qualities of individual learners.

MAKING THE EVALUATION QUESTIONNAIRES ACCESSIBLE

Although making a questionnaire or evaluation form (EF) accessible is not complicated, it is worth noting that even sighted people who are familiar with working with PSL can overlook how much sight is relied upon to scan ahead and compare possible responses in order to decide how to answer. It is also worth noting how simple changes can remove barriers resulting from impaired vision and facilitate a level of independence.

For ADLAB PRO relevant questionnaires were checked by PSL or by partners with the most experience in this area (principally RNIB and Soundfocus) to ensure their accessibility. In order to make the questionnaires easily read by PSL, it was necessary to remove graphics, such as the project logo, and to add a few explicit instructions that explained the layout. For example people were asked to type in a number rather than underline one on a pre-provided scale. Two examples are given below from the Multiplier Event (ME) evaluation questionnaire. For questions using a Likert scale, the original contained the rubric

To what extent do you agree with the following statements? Please use a scale of 1-5, where: 1 = disagree strongly; 2 = disagree; 3 = neither agree nor disagree; 4 = agree; 5 = strongly agree: 1 2 3 4 5.

This relies on respondents looking ahead to infer how to respond (by circling or underlining a number). The accessible version required an explicit instruction: “Enter the number at the end of each statement, after the colon.” Similarly, for questions offering multiple choice-style answers, more explicit information about layout is used to alert PSL to the fact that optional answers are available.

Original

The academic level of the content was (please choose one): too difficult/too simple/about right

Accessible version:

The next two questions give you a choice of 3 possible answers. Please choose the one that applies:

The academic level of the content was: too difficult/too simple/about right.

EVALUATING THE EVALUATION

Once the evaluation tool has been launched its success can be evaluated by reach and response rate (Fanning, 2005). Pinsonneault and Kraemer (1993, p.94) suggest that a response rate lower than 51% “is considered inadequate in the social sciences.” It should be noted that response rate should be calculated not only for the questionnaire as a whole but also for individual questions. Any systematic failure to address particular questions suggests an error of formatting or design, rather than a general problem with reach.

ADLAB PRO: Evaluation of Intellectual Outputs

EVALUATION OF IO1

This chapter demonstrates how the methods detailed above were applied with regard to the evaluation of the early outputs of the project and how these evaluations in turn fed into the development of the training materials created in IO4 (chapter 3). Outputs of ADLAB PRO are presented in the order they were evaluated so that the evolution of the evaluation and creation processes can be followed.

ADLAB PRO identified the problem that there are still few professional figures working in the field of audio description (AD) that they are often untrained and their skills are not clearly defined. The aim of ADLAB PRO is to fill this gap by identifying those skills (IO1 and IO2), defining a curriculum (IO3) and ensuring professional Europe-wide implementation of AD in all cultural and media sectors through the creation of online training materials (IO4). The successful completion of these steps was evaluated (IO5) before accreditation was awarded using ECTS and ECVETS (IO6). Each IO was evaluated as the following explains.

An internal evaluation process for each IO was carried out by the Advisory Board (AB) and by one of the Partners as allocated by the Quality Manager (QM). An EF was created with eight quality indicators in addition to an open field for other comments as shown in Table 1.

QI 1	The IO was innovative and advanced understanding of the field
QI 2	Showed rigorous methodology and analysis
QI 3	Contributed to the aims of ADLAB PRO
QI 4	The research was robust (good sample size; appropriate references etc.)
QI 5	The report is well structured
QI 6	The writing was clear and of an appropriate academic standard
QI 7	The IO was completed within a reasonable timescale
QI 8	The report covers all aspects indicated in and required by the original specifications
	Other

Table 1: Quality indicators for IO1

The evaluating partner for IO1 (UA) awarded each QI a mark out of 10, creating a total score of 78/80. Marks were dropped for QI4 and QI7; qualitative comments stated that it was "rigorous and sound: based on relevant literature and ample scientific experience. The limits of the sample size were due to the limited availability of AD courses." In the other comments section UA concluded it was "excellent work".

Recommendations from the AB were as follows:

- Learning by doing should remain a focus even after finishing a course.
- Peer review and discussion should be a part of the ADLAB PRO ongoing training modules.
- Ideally, it would be a platform that enables and stimulates the continuous exchange of knowledge and best practices between professionals throughout member states.
- Some theoretical framework on competences would help the project reach more scientific maturity and exert a long-lasting impact on the discipline.
- The report does mention the social-constructive teaching model and project-based learning (Kiraly 2000), which are very important in contemporary translator (and – by extension – audio describer) training, but the question of audio describer competences (skills?) is only mentioned in passing. Given the high-profile of the project and a great intellectual potential of its research team, I would suggest developing a competence framework, following the example of the EMT or PACTE models in translation studies. However, I believe that at this early stage of the project, there is still time to address this issue in more depth, possibly in the next IO when defining the profile of the AD expert.

Responding to the Recommendations

As IO2 was a consultative exercise (a survey with describers, providers and PSL), the recommendations were taken up in IOs 3, 4 and 5 (cf. respective reports on www.adlabroject.eu). It will be seen from the IO4 Final Report that practical tasks (learning by doing) feature in the training materials (tasks). Peer review and discussion are intrinsic to the evaluation process (IO5) described in this report. The project website (www.adlabroject.eu) is "a platform that enables and stimulates the continuous exchange of knowledge and best practices between professionals" not only throughout member states, but also globally. The theoretical framework on competences was addressed in depth in IO3. As stated in the IO3 Final Report:

> the research carried out in IOs 1 and 2 provided invaluable and very concrete additional and complementary input. Currently, courses in AD are taught inside and outside academia, for a varying number of hours, but they do not differ fundamentally in terms of the knowledge and skills they teach, the main distinction being the size of learner groups. The trainers are often practitioners who do research or practisearchers (Gile, 1994) and they make use of varied teaching/learning methods, as identified by Laurillard (2012) among others, which appear to yield good results (ADLAB PRO, 2017a, p 34). This was therefore taken into account in IO3. More than half of current academic courses are offered at master's level, with other courses spread evenly over bachelor programmes and postgraduate programmes. Non-academic courses take the shape of in-house training and workshops. This obviously varied picture fed not only into the modular structure of the course but also into the way the modules can be combined, the different levels at which they can be taught, and the suggested entry requirements. The different backgrounds of the practisearchers offering training in AD also inspired the design of manageable and adaptable modules.

EVALUATION OF IO2

IO2 was evaluated in the same way by UAM, although QI5 and QI8 were missing. This resulted in a total score of 58/60. Marks were dropped for QI 7 and QI 4. The qualitative comments explained "Small delays were experienced due to a variety of factors (technical issues with the questionnaire form, etc.)" and "183 complete responses to the questionnaire – an impressive data pool. The bibliography includes all major publications relevant to the scope of the report. Limitations of the study are clearly delineated. The geographical distribution of responses is not even." The overall impact of the report was summarized as: "An important and valid contribution towards the definition of an audio describing professional. "

Recommendations from the AB were as follows:

- Doing upcoming questionnaires also in these two languages (French and German).
- I felt the need for some summary and highlighting from time to time and especially in the end when my head was a little dizzy.
- In the chapter on quality – that it may be useful to have an analysis by country for some of the answers.
- I believe it would be interesting to explore the extent to which AD training can be considered "generic", i.e., its fundamentals apply to all formats in which it is practiced (performing arts/museums/media). (JS)
- Such a questionnaire – in my view – should not be the only way to build the project/course on. Therefore, I believe the project should address the issues of learning, acquiring skills and developing competence, focusing on the specificities of AD. What are the prerequisites (abilities?)

necessary to become an audio describer? Are there any? For instance, in the case of the translator, it would be fluency in two languages. It would be helpful to see how these abilities develop into competences and then into professional expertise. Is there a progression path envisaged in the AD training course? (AS)

- I was really interested to read about the professional status of audio describers and how it is perceived. This is a novel topic in its own right to be pursued in the future. I would recommend that the Project Team could publish the current findings as an academic paper, drawing on sociological aspects. (AS)

Responding to the Recommendations

These recommendations were discussed, however not all could be acted upon. As the evaluation was summative there was no opportunity to revisit IO2 and there were insufficient numbers to conduct an analysis by country for some of the answers, as highlighted in the Partner evaluation. In addition, there was no budget available to translate outside the languages of the project partners.

In the absence of an opportunity to act on all recommendations by retrospectively amending IO2, relevant comments were taken on board for IO3. In particular, IO3 "addressed the issues of learning, acquiring skills and developing competence, focusing on the specificities of AD." IO3 also specified "the prerequisites (abilities?) necessary to become an audio describer". The comment pertaining to "the extent to which AD training can be considered "generic", i.e., its fundamentals apply to all formats in which it is practiced (performing arts/museums/media)" is directly reflected in the structure of the curriculum developed by IO3. This established the main competences, specific subcompetences and Learning outcomes with an introductory module covering essential skills shared by all AD modes plus additional modules that were domain-specific. (Module 1: Introduction and transferable skills; Module 2: Screen AD; Module 3 Live events; Module 4: Static Arts and environments; Module 5 additional service and specific contexts; Module 6: Technological Issues, New developments).

EVALUATION OF IO3

Following the recommendations from the AB about IO1 – in particular "peer review and discussion should be a part of the ADLAB PRO ongoing training modules", it was felt that an external evaluation for the third IO would have the dual benefit of widening the pool of expertise and disseminating information about the project to key influencers and stakeholders. Consequently, a separate external evaluator was sourced for each of the 6 modules.

As explained above, it was decided that the key stakeholders in ADLAB PRO are trainers wishing to implement AD training in a variety of contexts. Partners were asked to nominate suitable individuals to act as evaluators and contact them to determine their general willingness to participate. This created a pool of 16 potential evaluators. However, it needed to be borne in mind that not all would be appropriate or available and that many evaluators would be needed for subsequent phases of the project (ultimately 30 key informants evaluated prototype materials in IO4). Utopian Voices (UV) selected from this pool, giving greatest consideration to the domain of expertise in AD in order to pair the most appropriate evaluator with each module. The aim was to balance other considerations such as geographical location and relationship to AD over the project as a whole.

In total 11 potential evaluators(two per module plus one spare) were contacted prior to the EF being sent. Research has long established that increasing the number of contacts between respondent and researcher

results in a higher return rate (e.g. Heberlein & Baumgartner, 1978; Goyder, 1987). The use of prenotification has also been shown to result in higher rates of return (Fan & Yan, 2010).
One PSL did not reply to the initial request. This resulted in 10 evaluators, representing a 91% response rate. They encompassed a diversity of stakeholders, all of whom were extremely familiar with AD and 7 of whom were active AD trainers. 6 of them were personally known to the researcher, meaning their qualifications did not need to be verified. Their experience is shown in Table 2.

ID	Familiar with AD	Known to the researcher	Relationship to AD	Mode of AD expertise	Trainer	From	Proposing partner	Sighted/PSL
001	✓	X	Provider	screen		NDL	UA	sighted
002	✓	X	Lecturer	screen	✓	NDL	UA	sighted
003	✓	X	Provider	Theatre		NDL	UA	sighted
004	✓	✓	Describer	Theatre	✓	UK	UV	sighted
005	✓	✓	Provider	Theatre & museums		UK	UV	sighted
006	✓	✓	Researcher	Theatre	✓	NDL	UA	sighted
007	✓	✓	Researcher	Museums		UK	UV	sighted
008	✓	✓	Multiple roles	screen	✓	GER	UA	sighted
009	✓	X	Provider	screen		NDL	UA	sighted
010	✓	✓	Describer	mixed	✓	NDL	UA	sighted

Table 2. Evaluators for Modules in IO3

All were asked to evaluate Module 1 as well as one domain-specific module relating to their area of expertise as shown in Table 3. They were not evenly distributed because it seemed more important that relevant domain specialisms of the evaluators were tapped and bearing in mind that Modules 5 & 6 were half-modules that do not deal with specific types of AD, but rather with additional skill sets audio describers need (or in some cases need to be aware of) in order to create appropriate descriptions.

.

Evaluator ID	Modules evaluated
001	2
002	2
003	3
004	3
005	3
006	4
007	4
008	5
009	6
010	6

Table 3. Evaluator by module

QIs for IO3

8 items were taken from the framework for IO3 employing indicators from Kennedy (2007):

- The module's competence framework encompasses the essential skills that should be learned by every describer.
- The module's learning outcomes provide a clear statement of what the learner should know, understand and be able to do as a result of completing the course.
- The module's learning outcomes are confusing.
- The module's learning outcomes are sufficiently broad not to limit learning.
- The module's learning outcomes cover all the essential things a describer needs to know in this AD context.
- The module caters to a range of learning styles.
- The demands the module makes on learners seem appropriate.
- Are there any learning outcomes or competences you would add or remove?

For each question evaluators were given a binary (Yes/No) closed response option coupled with a text box to provide a comment or evidence for their decision. A final text box allowed for any other comments to elicit more qualitative comments if desired or time allowed. This style of EF was created to ease the process for the evaluators, the limited number of response categories making it quick and easy to use, while the text boxes provided space to express feelings, in line with the findings of Preston and Colman (2000) whose

research showed that respondents found scales with fewer response categories easier and quicker to use but less good at allowing them to express their feelings adequately (this is discussed in section 8.2 of The Guide together with a detailed analysis of the results). The results of this evaluation are summarised below.

SUMMARY OF RESULTS

Quantitative data was derived from the number of respondents (dis)agreeing with the QIs. They were scored as follows: Yes=1; No=0; Y/N = 0.5; ?/??= 0.5; left blank = 0 giving a range of 0 – 6. The results are shown in Table 4.

Question Number	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Summed responses	5	5	1	6	3.5	3	6	0

Table 4. Quantitative Results from IO3 evaluation.

As expected these quantitative results are of limited use. Two of the questions (Q4 and Q7) received the maximum score of 6. This might be assumed to mean that Partners could rest assured that "The module's learning outcomes are sufficiently broad not to limit learning" and that "The demands the module makes on learners seem appropriate". However, even Yes did not come without reservations. Q3 received a low score of 1. This question (The module's learning outcomes are confusing) was worded negatively, making a low score desirable.

The qualitative data reflected the breadth of expertise of the evaluators and the differences between AD domains and between how AD is practised in different countries. For example, in response to Q1, one respondent commented:

> The module provides much more than what most AD'ers are actually doing (the AD'ers who work for the VRT[1] for instance, prepare their AD in 'Word' and leave the technical part to the technicians).

Respondents also made some very practical suggestions.

> I would recommend to separate [sic] the writing of an AD text from the voicing. Describers should know about the skills of voicing but not necessarily do it themselves.

Responding to the Evaluation

The qualitative data for each module was collated and circulated to the partner responsible for the relevant module. This allowed any concerns or suggestions raised by the evaluation process to be incorporated into the creation of the training materials as deemed necessary by the creator of that content. In addition, the following points were raised:

- Not all evaluators were aware that pre-requisite skills have been specified.
- Not all evaluators were aware that this is designed as a modular course.
- Not all evaluators were familiar with the language (jargon) despite being experts in AD.

[1] VRT – is the Flemish Radio and Television Broadcasting Organisation (Vlaamse Radio- en Televisieomroeporganisatie).

The evaluation process highlighted some deficiencies with IO3. These were rectified by the creation of a simplified course structure, which informed the design of IO4. (cf. IO4 Final Report, ADLAB PRO, 2019).

Evaluation of training materials (IO4)

The Evaluation Process In IO4

The evaluation process for IOs 1 – 3 provided lessons that were implemented in the process for IO4. This comprised multiple evaluations both formative and summative, using a mixed methods approach. Following recommendations by O'Cathain, Murphy and Nicholl, (2008) good reporting of a mixed methods approach requires that "the decisions behind this approach are explained". As mentioned above, this approach was justified by the desire to make use of specialist knowledge which required qualitative information, while minimising the demands on busy experts to help ensure their participation, for which quantitative responses were more appropriate.

The Purpose and Design

The purpose of the evaluation was to guarantee the quality of the training materials created, to ascertain their effectiveness, usability, and consequent longevity. The priority was to ensure that they would achieve their purpose of equipping students with the skills required by an audio describer. The next consideration was that the evaluation could be carried out within the available timeframe. In discussion with Partners, it was decided that evaluating a prototype[2] of each material was the most practical solution for formative evaluation.

Any revisions necessary could therefore be made before the final format was determined and then replicated in the multiple versions produced by Partners. This approach was a practical response to the sheer volume of materials produced (60 core videos + associated ppt. slides and transcriptions; over 100 additional videos; 6 reading lists; 196 tasks). Where possible both formative and summative evaluations were conducted. The various types of evaluation are shown in Table 5.

The detailed analyses are contained in *The Guide* and the results and conclusions are summarised below. It should be noted that the quality evaluations were carried out simultaneously with the accessibility checks. This was partly to ensure that everything could be produced within the timeframe but also to ensure that any revisions could be carried out in one go.

Material type	Formative Evaluation	Completed By	Supplementary evaluation	Summative evaluation
Core video (prototype)	External survey (mixed)	2 key informants	Accessibility (RNIB & SF) Technical & formatting checks UAB	Field testing (mixed)

[2] The term "prototype" has been adopted here in preference to "sample" which is used in other reports. This is to avoid confusion with the methodological meaning of "sample".

Introductory video (prototype)	External survey (mixed)	3 key informants	Accessibility (RNIB & SF) Technical checks UAB	Field testing (mixed)
Task (prototype)	External survey (mixed) Field testing (mixed)	3 key informants	Accessibility (RNIB & SF) formatting checks (UAB)	Field testing (mixed)
Reading lists (draft)	Internal (qual)	partners	6 key informants	Field testing (mixed)
Additional videos (final)		PSL	Technical checks (UAB)	External Focus group (qual)
Trainer's guide (prototype)	External survey (mixed)	3 key informants	Accessibility (RNIB & SF) formatting checks (UAB)	External trainers (mixed)
Task sample		Students /learners		Survey (mixed)
Materials (general)		Participants at ME5		External interviews (qual) Questionnaire at ME5 (mixed)
Materials (general)		Students /learners		Workshop (mixed)

Table 5. Methods of evaluation used in production of Training Materials (IO4)

Types of Evaluation

5 methods of evaluation were employed (surveys; field testing; shop window; semi-structured interviews and focus group). The types are summarized here and the actual implementations explained in detail below.

- **Surveys** were evaluation forms (EFs) of the type described for IO3 with different QIs as appropriate. These were sent to external evaluators (key informants) for peer review.
- **Field testing** involved incorporating selected materials into courses and workshops where possible, with evaluations collected from students and trainers. One of these was formative, using material that would later be incorporated into tasks and additional videos. Most were summative using opportunity samples of learners participating in courses/workshops that took place after the materials were ready and before the end of the project. The term "course" has been used to describe academic training whereas the term "workshop" has been used to describe non-academic training of interested professionals. All were conducted at three of the partner universities (UA; UAM and UNITS) together with a course taught at University College London by the director of one of the SMEs (UV). In addition responses from professional audio describers were obtained from freelance audio describers working at RTV-SLO.

- A **"shop window"** evaluation was conducted with participants at the Multiplier Event (ME5) held in Barcelona in March 2019. These stakeholders were shown selected materials and asked to complete questionnaires, with no opportunity to explore the materials in depth.
- **Semi-structured interviews** were conducted with selected participants at ME5.
- **A focus group** was organised with PSL at the RNIB in London.

The methodologies are presented in more detail and a summary of the results by material type before a general discussion. This will consider the results in terms of the evaluation of the IO4 training materials in particular. Lessons to be learned in the evaluation of training materials in general are considered in *The Guide*.

EVALUATION METHODOLOGIES

Field testing (student evaluations) UNITS

An Evaluation session with students took place on March 14, 2019 at the University of Trieste, at the Department of Legal, Language, Translation and Interpreting Studies, Section of Studies in Modern Languages for Interpreters and Translators (SSLMIT).

The following materials were shown and evaluated:

- 1 Core Video (Module 4 Unit 3: AD for static arts) https://ddd.uab.cat/record/202471
- 1 Additional Video (Module 4 Unit 4 (Strategies): AD of churches)

Both videos were presented in English without subtitles.

The participants were 119 Italian students studying Theory and history of translation. They were not AD students and had a very basic knowledge of the subject, which had been introduced to them during a 2-hour lecture. The session was conducted in Italian. After a brief introduction to the ADLAB PRO project, students were asked to watch a Core Video and to complete a Multiple Choice task followed by the evaluation form. Then, they were asked to watch an Additional Video and to fill in the evaluation form. The detailed results are analysed and presented in The Guide. A summary of the results is presented by material type in sections below.

Field testing (student evaluations) UAM

The Polish evaluation took place at the Faculty of English, Adam Mickiewicz University in Poznan during a course called 'Audio Description'. The course is an optional module for one semester in year two of a two-year specialisation in translation at Master's level. Since it is an elective course, the number of students is low (eight students had signed up for the course in question). As two were absent, six students took part in the evaluation. All were aged 18 – 49; all were from Europe and none spoke English as their first language. As for familiarity with AD, responses ranged from 1 "completely new" to 5 "familiar". This was probably due to the fact that some students had taken a course 'Introduction to Audiovisual Translation' two years before. This was taught by the same lecturer, where AD was discussed. Other students were completely new to AD. The lecturer is Polish and an experienced trainer in translation. The languages of instruction were PL and EN.

The following materials were evaluated:

- 1 Core Video (Module 2 Unit 2: Process) https://ddd.uab.cat/record/201952
- 1 Additional Video (Module 2 Unit 2: Workflow 01)
- 1 Task (Module 2 Unit 2: Prioritising information)

The Core Video and the Additional Video were presented and then discussed. The Task was done in Polish. The detailed results are analysed and presented in The Guide. A summary of the results is presented by material type in sections below.

Field testing: (learner evaluations) OPEN - UA & UV

A two-day workshop on AD for live events was held by OPEN the "Expertise Centre for Accessible Media and Culture" at the University of Antwerp 6th – 7th Nov 2018. 6 people attended, all working in theatre. The workshop was taught by Aline Remael, Nina Reviers and Louise Fryer, with input from Hanne Roofthooft, a PhD student at UA. The languages of the workshop were EN and NL. One completed core video was shown (1_CV_M3_U7). However, as this workshop took place before all the training materials were complete, it presented an opportunity for the formative evaluation of two power points that would be used for core videos in Module 3: PPT_M3_U4_UV_audio_introductions.pptx and PPT_M3_U4_UV_evaluation.pptx. In addition examples of clips of live events were shown that would be used for AV_M3_U1_1 as well as a draft task on evaluation for Module 3.

Field testing: (task evaluations) UCL - UV

A prototype task was evaluated by 9 students at University College London. The students represented an opportunity sample as they were studying AD as part of a year-long Master's degree in AVT. The demographic characteristics of the participants were as follows: m=3; f=6; 4 were native English speakers, 2 spoke another European language (Italian and French respectively) and 3 spoke Mandarin Chinese as their mother tongue. The course was taught in EN by the researcher. The whole task was carried out in EN, after the students had drafted their first AD script (also in EN). Although they had attended a couple of lectures introducing them to AD and the AD audience, the students may be thought of as beginners. The detailed results are analysed and presented in *The Guide*. A summary of the results is presented in section below.

Field testing: (professional evaluations) RTV-SLO

Four experienced audio describers at RTV Slovenia were asked to go through a selection of training materials (2 Core videos; 4 Additional videos; the Reading list for each module; 6 tasks and a trainer's guide) and fill in the evaluation forms. They were all familiar with the ADLAB PRO project so they did this individually. They were sent the materials so that they had time to go through them and to analyse them thoroughly without time pressure. All 4 participants are practicing describers, 1 is also an academic/researcher. They are freelancers at RTV-Slovenia. They are all native Slovene speakers who speak EN. They had some of the training in EN, and watched the materials in EN without problems. Three of them are in the age range 18-49, one 50-64. All of them are very familiar with AD. The partner who organized this evaluation (a journalist analyst specialist at the Department for Accessibility at RTV-SLO) commented: The respondents of the questionnaire are all experienced audio describers, one of them is also doing a PhD in accessibility – mainly in AD at University of Primorska in Slovenia, two of them are also teachers of pronunciation for the radio. All of them are writers of AD and also voice talents. They are familiar with the topic and they have done some

training but not in an academic way as in Slovenia there is (yet) no such programme at the university. The detailed results are analysed and presented in *The Guide*. A summary of the results is presented in section 3.2.5.2 below.

Shop window Evaluation (ME5): (stakeholder evaluations) UAB

On 21 March 2019 Multiplier Event 5 took place in Barcelona. Its aim was to present and evaluate ADLAB PRO training materials, and discuss AD training. The event included a general overview by project leader Elisa Perego (UNITS) and 5 invited presentations by trainers from different backgrounds and countries. During the event there were two evaluation sessions: the first session was led by Anna Matamala who introduced the rationale for IO4, the typology of materials and their main features. Examples of: (a) reading lists, (b) trainer's guides, and (c) an introductory video were shown. In the second session, led by Carme Mangiron and Anna Jankowska, the materials presented in more depth were: (a) core videos, (b) tasks, and (c) additional videos. Data was obtained from a questionnaire, which was made available to the ME participants both online and in hard copy. The detailed results are analysed and presented in *The Guide* (a copy of the questionnaire is included in the Appendix of *The Guide*). A summary of the results is presented by material type in sections below.

Questionnaires were received from 68 participants, 32 from session 1 and 36 from session 2. Most responses for Session 2 were obtained via paper copies, although 4 participants replied using the online survey. Their relationship to AD is shown in table 6. 40% of participants described themselves as having multiple links to AD. Of those with only one link, (22%) described themselves as an academic/researcher followed by student (14%) and practicing describer (8%). In the open field for "Other", they specified "partner", "artist using AD", "blind/consultant/software" or left the field empty. The majority came from Europe (29 or 91%) and one participant came from each of the following continents: Australia, Asia, North America. These results are shown in table 7. Over 90% (29) of respondents did not speak English as their first language.

	N	%
Practicing describer	3	8%
AD user	1	3%
Academic/researcher	8	22%
AD Tutor/Teacher/Lecturer	1	3%
Provider of AD content	3	8%
Student	5	14%
Other (specify)	2	6%
Multiple	14	39%

Table 6. Question 1: What is your relationship to AD?

Africa	0	0%
Asia	2	5.5%
Australia	2	5.5%
Europe	30	83.3%
North America	1	2.7%
South America	1	2.7%

Table 7. Participants' region of origin

Focus Group: (Evaluations of PSL) RNIB and UV

Given that the knowledge of the needs of visually impaired people was found to be one of the most important skills needed by an audio describer (IO1; IO2) some of the core and additional videos produced for IO4 contained information about the following: the heterogeneity of the blind audience (CV_M1_U6: https://ddd.uab.cat/record/200115); information about touch tours (AV_M3_U5_1) and how to guide PSL (AV_M3_U5_4) and what PSL found most helpful and least helpful when they were being guided (AV_M3_U5_3 and AV_M3_U5_2) (all M3 AVs can be retrieved from here: https://ddd.uab.cat/record/202312).

After each video, participants were invited to comment. A rough transcription was initially made using an online app called Speedscriber. This output was refined by the researcher. A full transcription is provided in the Appendix of The Guide. Selected comments referring to the Core Video are presented in the findings below. Comments referring to the Additional Videos are presented in section below.

5 people attended. All were either employees of RNIB, or worked there as a volunteer. All but one was female. Two of the five identified themselves as trainers, giving talks to community groups and to staff at Transport for London and Transport for All. In fact most were engaged in this type of advocacy. Having enjoyed a cup of tea and a slice of cake, participants signed a consent form which was read aloud. They were asked to state whether they were blind or partially sighted and to rate their familiarity with audio description (AD) using a scale of 1 –5. Their answers ranged from 2 – 5. In addition, the group convenor attended as an observer. This was Sonali Rai, who is the RNIB's representative in ADLAB PRO. Louise Fryer from UV conducted the discussion. Heather Temple (UV) also attended to facilitate the recording.

Semi-Structured Interviews (ME5)(Evaluations by Key Informants) UV

In order to supplement the ME5 questionnaires with more in-depth responses, 9 people (25% of those attending) were asked to undertake a short interview immediately after the 2nd session to give their evaluation of the materials presented. The interviews were carried out by UV and an audio recording made. This recording was later transcribed using the Speedscriber application and then refined by the researcher. The complete transcription is contained in the Appendix of The Guide.

The interviewees are listed in Table 8. They were selected for their experience as trainers – being the primary target audience for the materials produced by ADLAB PRO – and also for their global reach. 3 interviewees

are based in Europe; 3 in N. America; 2 in Australasia and 1 in Asia. All were highly familiar with audio description (AD). 5 of the 9 rated their familiarity with AD at 5/5, with the other four ranging from 3.5 – 6 (!). Having signed a consent form, each interviewee answered a series of questions about the materials they had been shown. Ethical approval was given by the University of Trieste. UV used Nvivo software to code and identify common themes, which are reported in full in The Guide and summarized in the relevant section below.

ID no	Region of origin	Self-rated familiarity with AD (1-5 scale)	Relationship to AD
Interviewee 01	Europe	5	editor for AD production
Interviewee 02	Australia	5	audio describer executive director of a non-profit AD service in mostly live events and visual art
Interviewee 03	Europe	5	audio describer University teacher of AD for a University masters course in AVT
Interviewee 04	North America	5	audio describer and trainer
Interviewee 05	Australia	4.5	audio describer and researcher
Interviewee 06	Europe	4	AD researcher and trainer
Interviewee 07	North America	6	Longstanding audio describer and trainer
Interviewee 08	North America	3	Audio describer "by marriage"
Interviewee 09	Asia	5	audio description trainer and practitioner

Table 8 Interviewees post ME5

SUMMARY OF RESULTS BY MATERIAL TYPE

- **Core videos: Prototype**

The first type of training material to be evaluated was a prototype core video produced by UAB. This was sent to 2 external evaluators. Both were highly regarded AD trainers in academic environments, one from Poland and one from Spain. More information is given in Table 9. Both evaluators responded positively to the request. A mixed methods approach was taken, again with an emphasis on the qualitative. The timing was dictated by the production timetable produced for IO4 by UAB, with a small window available between the prototype being ready and the definitive example needing to be produced in time for multiple versions to be created by partners. The location of the evaluation /style of distribution was remote. A similar process was adopted as for the IO3 evaluation.

ID	Familiar with AD	Known to the researcher	Relationship to AD	Mode of AD expertise	Trainer	From	Proposing partner	Sighted /PSL
E01	✓	✓	lecturer/researcher	screen	✓	ESP	UAB	sighted
E02	✓	✓	lecturer/researcher	screen	✓	POL	UAM	sighted

Table 9. Evaluators of the core videos.

In broad terms, it was decided to test the acceptability and usability of the training materials. The evaluation form is included in the Appendix (4.2) of The Guide. The specific quality indicators are given in Table 10. They were designed in line with the aims of the project and include UX indicators such as engagement and attention. Although these are psychological indicators, they are known to be linked to student success (Christenson, Reschly, & Wylie, 2012) and correspond to the category of reaction proposed by Kirkpatrick (1959).

	Quality Indicator
QI01	The video makes a useful contribution to understanding the practice of AD.
QI02	The video gives a good overview of the module to students of AD.
QI03	The video held my attention.
QI04	The video was succinct
QI05	The audio (voiceover) was engaging
QI06	If I were running a training course on AD, I would include this video (please give your reasons).
QI07	The duration of the video was about right.
QI08	The video is well structured.
	Other comments

Table 10. Quality indicators for core videos.

Results (Core video, prototype)

The prototype core video was awarded 59/70 and 67/70 (mean = 63). The qualitative responses of the evaluators were very detailed. They were critical in the sense of providing a thorough critique and went into considerable detail e.g. "I don't think full stops are necessary at the end of lines in bullet points if they are not full sentences." Negative comments included concerns over the pace of the spoken delivery; that the visuals were not very engaging and the lack of examples. On the positive side, both evaluators agreed that the structure was perfect and that they would include the video if they were running a course. One commented "Very good material. My comments are just suggestions that may help improving it, but the video is great as it is and I would use it in my lessons. No doubt. Great job!"

Responding to the Analysis

It was agreed that UAB should adopt the evaluators' suggestions or justify why they had not done so. For example copyright restrictions would make it difficult to act on the suggestion: "In min. 2 when talking about the examples, maybe add a screenshot of the Swedish TV or a picture from Inglorious Basterds?

By contrast this suggestion

> I would try to include a couple of short video examples (10-15 seconds?). The best place may be the slide dealing with "voice-over effect" and "dubbing effect". It'd be great to have the same clip with both effects. Another option would be presenting a single video at the very beginning showing an example of what AST is. In my lessons, some students don't understand the concept until they are shown a video with AST.

was incorporated into an Additional Video produced for the module. This had already been planned but the evaluation reinforced the need. The tight constraints around the format of the core videos e.g. that they be built around a ppt. presentation and their duration be 5 minutes (max), meant that additional videos provided a better medium to act on suggestions.

- Core videos: Final examples

After the formative evaluations of the prototype, a selection (5/60) of final core videos were subject to numerous summative evaluations: the "shop window" evaluation at ME5 via the evaluation questionnaire and also the interviews carried out with selected participants; the focus group at RNIB and also evaluations with learners and students at UNITS, UAM, UA and professionals at RTV-SLO. The results are summarised here (with detailed analysis in The Guide) to give a thorough indication of the reception of the core videos in terms of Kirkpatrick's categories of reaction and learning.

Core videos, Final Examples: Reaction (ME5)

Summary of results

The quantitative data produced excellent results. The most common response (mode) to all the positive QIs was the top value of 5, meaning most respondents strongly agreed that they found the core videos interesting, well-structured and easy to understand. Most also strongly agreed that the core videos increased their understanding of AD. This was more so for students and those with multiple connections to AD, than for practising describers. Most of the respondents strongly disagreed that they found the videos confusing. The qualitative results showed appreciation for the simplicity, succinctness and clarity of the videos. Fewer participants specified the things they liked least. Those that did picked out that the videos were a bit dull in terms of visuals and that the delivery of the presentation was a bit fast. One person was concerned by the use of a linguistic register that was overly academic.

Core videos, Final Examples: Field testing (student evaluations) UNITS

Summary of Results

In terms of reaction, the Core Videos received positive scores from the Italian students most of whom agreed they were clearly presented and effectively organized. The Core Videos also received positive scores in terms of learning with the students mostly agreeing that they were easy to understand, provided appropriate guidance on the topic, increased knowledge of the topic and developed skills in this subject, In total 94% of those who responded to the question would recommend the video. 81.5% of students answered all questions correctly in the Multiple Choice and rather than a normal spread, the results were clustered towards the upper end. This could be interpreted in two ways - either the video is an excellent teaching tool, or that the multiple choice is too easy. It is likely that Q4 is too easy, as it was answered with a 100% success rate. Bivariate correlations showed that achievement in the MC was significantly positively linked to agreeing that the contents of the core video:

- provided appropriate guidance on the topic ($R = .348$, $p<.001$)
- increased my skills ($R = .349$, $p<.001$)
- developed my skills ($R = .287$, $p=.002$)

The correlations increase confidence that, for the students, the core video was an excellent teaching tool. Of the students who got full marks most (59) found the contents neither easy nor difficult to understand. Compared with the participants at the ME, the Italian students considered it slightly less easy to understand. This is not surprising as the Italian students had only just been acquainted with AD. Consequently, it can be concluded that CV_M4U3 is a successful teaching tool.

The qualitative results are remarkably consistent with those of the interviewees and participants in the ME5 in Barcelona, although there were some contradictory opinions as to whether the pace was good or too fast. The Italian students were not asked to evaluate the amount of mental effort they put into following the Core Videos. Instead, this measure was used in evaluations with a small cohort of students at the Adam Mickiewicz University in Poznan and with AD professionals at RTV-SLO.

Core videos, Final Examples: Field testing (student evaluations) UAM

Summary of Results

This was a small cohort but the results were surprisingly consistent with those of previous field tests described above, in evaluating comprehension of the core video and the ease with which the students were able to follow it. In the first case responses ranged from “quite good” to “very good”, with the average being “good”. Results for ease of following ranged from “fairly” to “very”, with the average response being “quite”. The level of difficulty was evaluated from “quite easy” to “easy”. Most of the respondents found the pace of the video “neither fast nor slow”.

For the other tested aspects there was much more variety in the answers. For example, respondents were able to devote “little” to “a lot of” attention to the video (the average being “neither little nor a lot”), some of them found retaining the general information to be “reasonably” easy, others “extremely easy” (on average: “quite easy”). Even more variety was found with the follow-up question about retaining specific information. Here responses ranged from “not at all” to “extremely”. When it comes to the interest in the video, responses ranged from “boring” to “exciting” (the average being: “neither boring nor exciting”). Similar

discrepancies were found for the question about the mental effort invested in the video (scores were evenly spread from 2 to 8 on a 9-point scale, where 1 = "minimal effort" and 9 = "extreme effort").
Overall the Core Video was evaluated positively, it was quite easily comprehended and did not pose a significant cognitive challenge to the majority of students. Of more concern, some did not find it interesting. An issue to consider here is the make-up of the group – as already mentioned, some students were already familiar with AD and chances are the materials were not properly suited to their level. The results indicate that video was more positively assessed by beginners. This would be consistent with the positive response given by the students in Trieste who were all new to AD.

Core videos, Final Examples: (professional evaluations) RTV-SLO

Summary of Results

2 Core Videos were tested:

- CV_M3_U3. This is from the unit for AD of live events and is designed to help learners to select what to describe in a live performance. (https://ddd.uab.cat/record/202310)
- CV_M6_U2 concerns the technology used to deliver AD. (https://ddd.uab.cat/record/202681)

This small group of professional describers all agreed or strongly agreed that contents of the Core Videos are clearly presented; effectively organised; easy to understand; increased their knowledge of the topic and developed their skills on this subject. Their reaction to the content and structure is in line with others described above. However, their response to the "learning" aspect is more surprising as they are already professionals working in the field of TV AD. One explanation is that the videos deal with AD domains with which the professionals were unfamiliar. CV_M3_U3 deals with the differences between AD for live events and AD for screen, and CVM6_U2 addresses technical delivery of screen AD. All the respondents are creators of AD content rather than being involved in the technical side of AD. Although this cohort is extremely small it is useful because it suggests the Core Videos can be valuable in a professional as well as an academic setting.

Limitations

The major limitation with this evaluation is the small group size. For example, it is hard to remain anonymous. Both the RTV_SLO professionals and the students at UAM might have been tempted towards researcher bias, the students wanting to please their teacher, the professionals wanting to please their superiors, who as ADLAB PRO Partners organised this evaluation and who are also responsible for offering the freelancers work. In order to overcome the group size limitation, the two groups were combined for further analysis. The advantage of combining the two groups is that the results become more reliable while also negating some of the extraneous influences such as teacher input and environment that are known to affect the impact of training materials on learner attainment. From this we can be quietly confident that, with a mode of 5, the mental effort required by the Core Videos they watched was neither too little nor too much. Although it is not possible to generalise from this about all the Core Videos, a common pattern is beginning to develop. This provides justification for the decision to subject a prototype to formative evaluation, creating a model that all final versions followed.

Core videos, Final Examples: Focus Group

Summary of Results

The focus group comprised participants with varying experience of AD. However, they were all experts in the field of sight loss, both from personal experience and in their professional roles at the RNIB as advocates for PSL and as trainers of sighted people who need to learn how to interact with and assist PSL. The content of the video introducing the Target Audience for AD received a mixed response. One person thought it "really, really good" but another disliked the emphasis on AD for all, feeling "some more emphasis could be put on blind people" Participants also had reservations over the explanation, presented in the video, that the purpose of AD was to help PSL understand a TV programme and also over the language used in the video which was not always politically correct. The group did not shy away from giving critical feedback but it was clear that videos touched them enough to provoke discussions relating to issues of concern, namely the audience for AD and the language used to describe disability. The amount of information in the video and its contribution to raising awareness of the needs of PSL were also appreciated.

Core Videos were also evaluated as part of responses to training materials in general and will be revisited in section 3.2.8. For now, attention turns to the evaluations of other types of training materials before their strengths and weaknesses are discussed.

- Reading Lists

Internal evaluation (Reading lists: formative)

Reading lists underwent an initial internal evaluation as shown in Table 11. Partners were given a deadline to comment on all lists after which the Partner responsible for each module refined the list. Subsequently, Partners were asked to comment on the refined version of a specific list as shown in Table 20. No evaluation form was produced and no quality indicators were given. Instead Partners were simply invited to comment on whether or not references were appropriate and invited to add any they felt may have been overlooked.

Module number	Topic	Reading list produced by	Reading list assessed by
Module 1	Intro to AD	UA	UNITS
Module 2	Scriptwriting for recorded AD	UAM	UAB
Module 3	(semi) live AD for dynamic performances and events	UV	UA
Module 4	(semi) live AD for static arts and environments	UNITS	UV
Module 5	Additional services	UAB	UAM
Module 6	Additional technical issues, developments and change	UAB	UAM

Table 11. Internal evaluation of Reading Lists schedule

External Evaluation (Reading lists: formative)

Following this internal evaluation, an external evaluation took place. This was to avoid the project becoming too inward looking, and for the purposes of openness, dissemination and extending its reach. UV matched each list with an appropriate evaluator (Table 12). Of the six evaluators approached one was unable to complete the evaluation due to illness (003). Consequently a seventh was recruited (007). Of the final six, three of the evaluators are based in the UK, one in Italy, one in Canada and one in Australia. All are actively engaged in AD research and are Professors or have a doctorate in a discipline related to Audiovisual

Translation (AVT). In addition to the quantitative section, evaluators were asked to list any references that they would remove and any missing references that they felt should have been included.

ID	Known to the researcher	Relationship to AD	Domain of AD expertise	From	Sighted /PSL
001	✓	Lecturer/researcher	live events	UK	sighted
002		Professor	Screen/live events	CAN	sighted
003	✓	Lecturer/researcher	live events	UK	sighted
004	✓	Researcher	Museums	UK	sighted
005	✓	Professor	screen	AUS	sighted
006	✓	Researcher	AD sound	UK	sighted
007	✓	Professor	Screen/live events	ITA	sighted

Table 12. Evaluators for Reading lists.

Method (Reading Lists, external, key informants)

The evaluation form completed by each evaluator comprised 5 statements to which evaluators were asked to respond using a 1-5 Likert scale (from 1: strongly disagree – 5: strongly agree). One question "The references were in APA format" became redundant as UAB had taken care to ensure that all the lists were correctly and consistently formatted, so responses to that question are not reported. The detailed analysis is contained in The Guide and a summary given below.

Summary of Results

With regard to the quantitative results, the scores are high (and therefore positive) being consistently above 4.5 for all indicators, suggesting the references are appropriate, relevant and sufficient. However, they only give a general indication as each evaluator was responding to a different stimulus. Qualitative data was particularly important with multiple suggestions of references to be included. As this was a formative evaluation, the evaluators' comments raised the following issues for partners to discuss:

- Should all evaluators' suggestions be adopted?
- Are there too many references per list?
- Should some be identified as essential, others as less so?
- Should information be included in the trainer's guide as to how the reading lists are to be used?
- Are all lists too Eurocentric? (N.B. M2 was the only list evaluated by an expert from N. America; the only other expert not based in the EU evaluated M5.)

Responding to the Evaluation

It was decided that either all evaluators' suggestions should be adopted, or the partner responsible for the relevant module should write a short justification of why not. Currently there are 10 references per unit. It was agreed that 10 should be retained, of which 5 should be highlighted as essential texts. All Modules should check they have included some non-EU references, although it is acknowledged that this will not be

possible, as so much research in AD has been conducted in Europe. It was also recognized that although the Reading List evaluated by the non-EU academic had only EU references, many of those she suggested were already represented in Reading Lists for other modules. Beyond the Reading Lists, the evaluators' comments affected other types of training material namely the content of the Tasks, which introduced a time estimate for required and recommended reading. The trainer's guide also gave an opportunity to explain more about the need for the localization of the materials that the trainers would be expected to implement themselves.

Reading Lists: Shop window (ME5)

Summary of Results

An example of a Reading List was shown at Session 1 of the ME5 in Barcelona and subject to the same "shop window" evaluation process described above. The results from 32 participants was good with all the positive QIs rated above 4.7 out of 5 meaning participants stated the lists were interesting, well-structured and easy to understand. Most participants strongly disagreed that they were confusing. As for the qualitative comments, there was disparity between those who thought the reading lists were "Too long"(UAB29) and those who liked the variety and richness and found them comprehensive e.g. "I liked the great number of references." (UAB25). It was interesting that two respondents praised "That they give 5 basic and 5 additional reading suggestions" (UAB26) and "The idea of choosing basic and crucial material for students" (UAB17) as this was a change introduced following the formative evaluation process described above. At this stage of the project some of the suggestions e.g. "Did you consider including a 3 line summary of the content for each publication?"(UAB005) were too late to act on. In any case some people already felt the reading lists were too long (UAB029).

- Introductory Video (prototype)

A prototype of an introductory video was created by RTV-SLO. After being commented on by partners (formative, internal evaluation) the video was sent for a formative evaluation to three external evaluators with an evaluation form (EF) containing 6 QIs evaluated on a 1-5 Likert scale:

1) The video was a helpful introduction to the module:
2) The video was well structured.
3) I found the video interesting.
4) I found the video confusing.
5) I found the video engaging.
6) The video made me want to find out more about this module and the ADLAB PRO course.

Evaluators were also asked to state what they liked best/ least about the video and to complete the sentence This video could be improved by… The EF and detailed results are contained in The Guide.

The evaluators had all previously been approached and asked if they would be willing to evaluate for ADLAB PRO. Two of the evaluators were from the UK and one was from Italy, all were highly familiar with AD. Two were based in academic settings, one described himself as an Audio Description Projects Manager and trains his own staff in AD. Their characteristics are shown in Table 13.

ID	Familiar with AD	Known to the researcher	Relationship to AD	Mode of AD expertise	Trainer	From	Proposing partner	Sighted /PSL
E01	✓	x	lecturer/researcher	screen	✓	IT	UNITS	sighted
E02	✓	✓	AD provider	screen	✓	UK	UV	sighted
E03	✓	x	lecturer/researcher/PhD student	mixed	✓	UK	UV	sighted

Table 13. Evaluators for the introductory video prototype.

Summary of Results

The video was evaluated by three high-profile persons: Two from academia and the third a trainer within the industry. Two were native English speakers and the third a native Spanish speaker, working in Italy. The quantitative results were not as impressive as for other material types but they were, nonetheless, good. All the positive indicators (that the video was well-structured; interesting, engaging and a helpful introduction to the module) were rated at 3.67 or above. A mean of 4.33 was achieved by two of the indicators (interesting, engaging) and also by the most significant indicator for the purpose of the introductory video and for the project overall: The video made me want to find out more about this module and the ADLAB PRO course.
As for the indicator phrased negatively (I found the video confusing), the mean was pleasingly low (m= 2).

The qualitative comments were also helpful. Two of the evaluators picked up on typographical and grammatical errors that had been missed during the internal evaluation. There was an inconsistent spelling (both "voice over" and "voice-over" are used) that had gone unnoticed by Partners. There was a discrepancy between the evaluators in terms of the audio. One liked "the warmth and softness in the Voiceover's tone and her clear speech", and another liked "The simple, straight-to-the-point style and friendly delivery" while the third "had to listen to the audio several times to catch the phrase: "Deals with additional services."
In terms of the video (as opposed to the audio), one evaluator suggested that "the way [the cartoon presenter] is dressed is not appropriate for this presentation." However, as the other comments on the visuals were positive ("I like the uncluttered style of the video with text and messages that are big and clear to read. I also like the unobtrusive background images."), it is suggested that they are left unchanged.

Responding to the Evaluation

Following the external evaluation, RTV-SLO amended the orthographical rendition of "voice over" to make it consistently "voice-over", in line with most of the references in the Reading list for Module 5.The phrase "Deals with additional services" was re-recorded. In line with the internal evaluation, a shortened version of the disclaimer was read out in the voice-over.

Introductory Video, Final Example: Shop window evaluation (ME5)

The amended introductory video and another for the whole project were among the items evaluated at the ME5 in Barcelona.

Summary of Results

The purpose of the introductory videos is to provide an enticing overview to each module, to encourage visitors to the ADLAB PRO website to find out more. Consequently, the "shop window" evaluation of ME5 was more suited to these videos than to any of the other types of IO4 material. Despite this, some of the qualitative comments indicate that some participants would have appreciated a longer look before giving an evaluation. Some of their concerns about the visuals echo those raised by the evaluators of the prototype, although this was not universal and there was even specific praise for the video being "visually, appealing and engaging". The quantitative results were very positive. The most common response (mode) was 5 for all QIs, except for finding the videos confusing, to which the most common response (mode) was 1. This was a desirable outcome (strongly disagree).

- Additional videos

Over 100 additional videos were created as part of IO4. They were extremely diverse and some made use of material originated elsewhere. Given such diversity, no prototype was created and no formative evaluation was carried out. The videos were subject to an internal assessment for technical quality carried out by UAB and selected videos were evaluated by participants at ME5 (session 2), field tested by students and learners at various courses/workshops, and by Slovene professionals. Qualitative evaluations came from semi-structured interviews at ME5 and PSL at the focus group organized at the RNIB. In total 10 additional videos were evaluated (10% of the total). The results from each evaluation are given below.

Additional videos (ME5)

3 Additional videos were shown at the ME (Session 2), as follows:

- M1, U6: The target audience of AD. This video shows interviews with PSL talking about their response to AD. It was created by RNIB and lasts 5 mins and 11" and can be accessed here: https://ddd.uab.cat/record/202686?ln=ca
- AV_M3_U2: Technical skills. This video explains some of the technical skills needed by a describer of live events. The video lasts 5 mins 13" and can be accessed here: https://ddd.uab.cat/record/202308?ln=ca
- AV_M5, U1: Audio subtitling. This video explains the concept of audio subtitling and gives some examples. The video lasts 3 mins 30" and can be accessed here: https://videosdigitals.uab.cat/almacen/downloads/461/11161.mp4

Summary of Results

The quantitative results were very positive. Of all the materials shown at ME5, the Additional Videos were rated the most highly. The mean average for every statement was above 4.7 on a 5-point Likert scale. The only exception is Statement 3, which was phrased differently and where the value (below 1.5) represents positive feedback. The things participants liked best about the Additional Videos was their diversity, that provided they personal experiences from professionals and users and gave different vantage points from AD users, audio describers, etc. They also included examples of real AD and "seem very entertaining too!" Furthermore "I got a bit emotional watching the first video – it's wonderful to see how much AD enhances people's lives." Negative comments included that trainers will need to watch them all when choosing what to use and that sometimes, they are too fast – "too much info into less time". One person commented "Why is the user experience/expectations an additional video!? This should be core material".

Of concern in terms of evaluation was the high number of non-responses to the qualitative questions, especially to what participants liked least (50% of participants did not complete this section). This means that the opinions of those who did respond are overrepresented. It is unclear whether there were more responses for "liked best" than "liked least" because respondents struggled to find anything to complain about or whether this was the result of fatigue or some other reason. It may be because they wanted to be supportive of the project or, because they could see the value of the materials overall, they were unwilling to criticise. Whatever the reason, this is the first time any comment had been made about engagement and emotional involvement in the training materials. As this response was unexpected it was not covered in the QIs assessed by the evaluation forms and shows the value of a mixed methods approach.

Additional videos: Student evaluations (UNITS)

Evaluation sessions with students took place at the University of Trieste, as reported above. One additional video was shown: AV_M4_U4 (Strategies): AD of churches. It was presented in English without subtitles. The students completed the EF assessing the videos on 6 QIs using a 1 – 5 Likert scale.

Summary of results

The Additional Video was highly rated with 70 of 117 students (58%) agreeing the contents were clearly presented and 62% that they were effectively organized. 49% understood the contents quite easily and 90% agreed or strongly agreed that they provided appropriate guidance on the topic. Overall, participants showed a more positive attitude to the Additional Video compared with the Core Video. Quantitative responses to the QIs evaluated on the Likert scale were broadly similar but 112 participants (99.12%) would recommend the Additional Video, compared with 107 (94%) who would recommend the Core Video. Seven students stated they would not recommend the Core Video. Only one would not recommend the Additional Video. Regarding qualitative answers, positive answers were similar for both types of video (both are considered, for example, to be easy to understand, to concern an interesting topic and to broaden knowledge). The pattern was repeated of negative answers being less frequent compared with positive ones.

Additional videos: Student evaluations (UAM)

Summary of results

The exact same tendencies in variation in responses where found for the Additional Video as for the Core Video. Most of the scores were broadly repeated, with modes being practically the same. This may be due to the fact that the students filled in the questionnaire at one sitting, and thus the answers for the Additional Video may have been influenced by those already given for the Core Video. As explained above, this small data set is of little value on its own consequently it was combined with the responses from the 4 audio describers at RTV-SLO. Despite this pair of small cohorts comprising very different types of learner (Polish students and Slovene audio describers) who were assessing a different video, the combined results show a surprising consistency with those of the larger groups (the Italian students and the participants at ME5) in expressing a preference for the Additional Videos over the Core Videos. This might be because the additional videos are more varied and didn't suffer the same constraints as the Core Videos in terms of timing, visuals and presentation style. It may well be that the basic theoretical information presented in the Core Videos was simply less exciting (and certainly less emotionally engaging) than the additional content.

Additional videos (focus group)

Four Additional Videos were shown to the focus group of PSL at the RNIB:

- AV_M3_U5_1.
- AV_M3_U5_2
- AV_M3_U5_3
- AV_M3_U5_4

Responses to these particular videos are discussed below:

Discussion

In general participants felt it was useful for students of AD to learn about touch tours and the importance of touch for some PSL but there was a lot of negative discussion around the portion of the video that showed a blind man at a touch tour as his narration left no room for the images to be described, as this comment illustrates

> I would say that for that bit I did get a little bit disengaged because I felt it needed more description. So I felt like he was getting the experience with touching it that I wasn't getting with listening to it (P04).

M3_U5_2 Helpful guiding

This video shows a number of PSL talking straight to camera about what they find most helpful when being guided. One participant suggested this could be improved by showing somebody demonstrating how to guide well. On the negative side, the expression of at times contradictory views by the PSL in the video was thought to be potentially confusing but this opinion was not universal

> I like how it showed each person talking, explaining their point of view, the guiding and what they like.
> Yes. But I also agree that they should have said at some point that you should ask the person first. (P05)

Generally this video was well received. Positive key words and phrases included: useful; relevant; very good; a lot of information; helpful.2 participants expressed a wish for it to be seen by a wider audience and the others agreed:

> And I just wish that more people could see it [this video]. Especially at the train stations. Yeah. You know it really does need to be out there. And even in families. My daughter's the worst guide. (P01)

> I personally really liked it and I'd like a copy of it. Yeah it was very good. (P02)

AVM3_U5_3 Bad Guiding

This video shows PSL talking about how they feel when they are guided badly. This video did not meet with instant approval.

> Well. I think that's going to freak sighted people out. I do. I think the first thing that's on that video is how dangerous it is if you get it wrong. And it kind of takes that thing away of "Look just take it easy" Just ask them what they need. I mean. It just makes it such a big deal. (P04)

Although some participants recognised it allowed for learning by being shown what not to do.

> This is a kind of you know like they don't do it perfectly kind of an interview. (P03)

M3_U5_4 Guiding

This was a didactic video, using a ppt. presentation, in the style of the Core Videos. It received universal approval when the Researcher asked if it was an acceptable way to talk about PSL.

Discussion

The focus group revealed potential users of ADLAB PRO materials that had not previously been considered, such as anyone concerned with PSL. Those mentioned by participants include local blind societies, workers who are public-facing such as railway staff and even friends and family members of PSL.

Not to be ignored was the negative feedback concerning a lack of engagement with the content of one video (M3_U5_1). This stemmed from minimal AD at certain points calling into question the accessibility of the video. However, it should be noted that this stemmed from the constraint that AD has to be woven around the existing soundtrack. At the point of concern, Trev, the narrator, is talking all the time, his words illustrated by footage of a touch tour. Trainers might like to use this as a point of discussion when showing this video to learners. The most positive aspect for ADLAB PRO was the unanimous interest expressed by participants for incorporating the videos into their own training programmes. This was a huge acclamation which demonstrates how well these videos fit not only with the aims of the project but also go beyond it into the realms of advocacy for PSL and training for sighted people in how best to assist PSL. This augurs well for the project's sustainability.

It should be noted that following a tweet about the reception of these videos, the researcher was contacted by a PSL who is the business and innovations manager at Vision West of England. Having requested and watched the video, he commented. "Really liked it!" and expressed an interest in developing training and standards around orientation and mobility training for PSL.

- **Trainer's Guides (prototype)**

The process for evaluating the trainer's guide used an evaluation form sent to 3 key informants who were asked to evaluate a prototype.

All the evaluators agreed to the request. 2 are from the UK and are highly experienced trainers and members of the Audio Description Association (ADA), the third is an audio describer and PhD student in Slovenia. Their characteristics are shown in table 14. The two from the UK mostly run vocational training courses while the Slovene evaluator is from academia and also a professional audio describer.

ID	Familiar with AD	Known to the researcher	Relationship to AD	Mode of AD expertise	Trainer	From	Proposing partner	Sighted /PSL
E01	✓	✓	Audio describer/trainer	stage	✓	UK	UV	sighted
E02	✓	✓	AD provider/describer/trainer	stage	✓	UK	UV	sighted
E03	✓	x	lecturer/researcher/PhD student/ describer	mixed	✓	SLO	RTV-SLO	sighted

Table 14 Evaluators for the prototype Trainer's Guide

Results

All three external evaluators agreed they had enough experience in teaching/training to assess this guide (mean = 4.3). Despite their different backgrounds their responses were broadly consistent. The quantitative results were good. A mean of 4.3 was achieved by all but one of the positive indicators (clear, informative and fit for purpose). A mean of 4 was achieved for the remaining indicator in response to "The trainer's guide would help me decide whether or not to use the ADLAB PRO training materials if I were teaching a relevant course." The mean for the negatively phrased indicator (The trainer's guide was confusing) was pleasingly low (m= 1.3).

The qualitative comments mostly supported the quantitative results. Only one evaluator made a concrete suggestion for improvement– that the LOs should have accompanying assessment criteria.

Responding to the Evaluation.

Partners considered the suggestion that UAB should add assessment criteria for each LO, but decided not to proceed, given that trainers using the materials would be doing so in different learning environments, to which different assessment criteria might apply.

Trainer's Guide Accessibility Evaluation

All training materials were subject to an internal accessibility evaluation. The one for the prototype trainer's guide was carried out by RNIB who reported:

> Add a line break between each content line e.g., Add a line break between "Introduction and Module Structure and Learning Outcomes. Reason: At the moment, screen reader is reading consecutive lines as one straight line e.g., Introduction and Module Structure 22 Learning Outcomes 53.
> Structure 2. The document isn't structured properly for screen reader users. List bullets must be used throughout the document but what we have now is a combination of different styles including list bullets and bullet points. E.g., right now, the following bullet icon is read as 'o' by the screen reader. So the sentence reads - O An additional video with dubbing examples. And so on... Honestly, I think this is just an oversight as the author has stuck to the accessible style in all other respects."

Responding to the Evaluation.

The guide was reformatted in line with RNIB's suggestions.

Trainer's guide (ME5)

A Trainer's Guide was introduced and shown at the ME5 in Barcelona. Participants evaluated it following the process outlined above.

Summary of Results

The quantitative data are very good, with means above 4.7 on a 5 point scale for the positive indicators suggesting that most participants agreed or strongly agreed the guides were interesting, well-structured and easy to understand. Most strongly disagreed they were confusing with a mean of 1.19. There was a high number of non-responses to the qualitative comments. Slightly fewer people articulated "what they liked least" compared with "what they liked best" and the responses of those who did reply were conflicting, with some people liking the wide range of topics and others feeling the lists were boring, being too long with too many details. It is unclear whether people find it easier to find points to praise rather than criticise because

the guides are good or because of the novelty of having resources available for trainers. One respondent expressed what they liked best: "that there is a guide for trainers"

- Tasks (prototype)

A prototype Task sheet was created by UAB, based around Module 6. It comprised a series of multiple choice (MC) questions based on the Core Video, followed by a more active task designed to engage students and develop their skills in the topic of concern. UV sent it to three external evaluators who teach AD in academic institutions together with the Core Video. The evaluators were from three EU countries, namely Poland, Spain and Italy. All are Professors or have a doctorate. The evaluators were asked either to use the task in class or to imagine doing so. The evaluation report also incorporated Partners' comments from an internal evaluation process, so that all feedback could be considered together.

The evaluation form was similar to that used for the Core Video. It comprised 8 statements to which evaluators were asked to respond yes or no with space to provide evidence or a comment. One statement offered three possible answers (too many; too few; about right). In addition the evaluators were asked to make an estimate of CL using a 1-9 Likert scale (from 1 = minimal effort – 9 = extreme effort).

Summary of results

The sample tasks prompted more criticism internally than externally. All the external evaluators agreed that the tasks are well structured and that their students would find the tasks interesting. Also the demands of the task appear appropriate in terms of cognitive load, which at m=6 is neither too easy nor too demanding. Perhaps more importantly all the external evaluators showed an appetite for more tasks like these from ADLAB PRO.

Responding to the Evaluation

Following the internal and external evaluation of this prototype task, the following changes were agreed:

- The number of questions should be reduced to 5
- The questions should be based on the core videos only
- The questions should have a single correct answer
- An attempt should be made to make the task description more succinct
- The time estimate should apply to the whole task

Tasks: prototype (student)

One problem with evaluating a prototype task is that only the format was replicable. Consequently another type of formative evaluation was conducted on a selected task for Module 3. This was evaluated by 9 masters students at University College London. The task required them to

1. Exchange your AD script with a colleague. Using the ADLAB PRO assessment sheets, evaluate each other's work.
2. Revise your script in line with the assessment. Write a short paragraph (max 500 words) outlining what changes (if any) you made in line with your colleague's suggestions and what you thought of the evaluation process.

Following the task, the students completed a task evaluation form, comprising 2 demographic questions (gender and mother tongue) and a response to 6 statements using a 1- 5 Likert scale. In addition they were

asked to estimate the amount of mental effort it took to complete the task and also to estimate how long the task took to complete.

Summary of results

Overall the task was received well. All students agreed or agreed strongly that the task improved their understanding of the needs of AD users and that it increased their understanding of AD. 6/9 students strongly disagreed that they found the task confusing and all either agreed or agreed strongly that the demands were appropriate. The task was deemed interesting by all the students, 4 of whom agreed with the statement "I found the task interesting" and 5 of whom agreed with it strongly. In terms of how much mental effort it took to complete the task responses ranged between = 6 – 8. The most common response was 7. The mean estimate of how long the task took to complete was 56.25 minutes

Discussion

The ability to give and receive criticism is one of the soft team-working skills whose importance was highlighted in IO1:

> "According to Kiraly and his emergentist model (2000, 2003, 2005), teaching translation should be based on situated learning and should develop transferable (soft) skills, i.e. such skills that are not closely linked with one particular profession but instead can be transferred to other jobs and workplaces." (ADLAB PRO 2017: 2).

IO2 discovered that Team-working skills were considered important or extremely important by 60% of existing describers and 73% of service providers, suggesting that this task is ecologically valid and would help students succeed in the workplace. It is also important in terms of time management that they are aware how long it takes to complete an AD script. This was also important for the project for the allocation of ECTS and ECVETS (IO6). All in all, this task was retained in the final set of tasks for M3 and is recommended to trainers running AD courses in the future.

Tasks: Final (student/professionals)

4 freelance audio describers at RTV-SLO evaluated Tasks from 2 units in Module 5 and one in Module 6:

- TK_M5_U2
- TK_M5_U3
- TK_M6_U1

In addition the 5 Polish students from UAM evaluated 1 Task in Module 2: TK_M2_U2

Their combined results are reported in Table 15 below

Do you think your comprehension of the instructions for each task was (very poor – very good)	Mode = 6 (good)
Mental effort	Mean = 5.6
How easily were you able to follow the task instructions? (not at all – extremely)	Mode = 4 (neither easily nor with difficulty)
Did you find the tasks… (very boring – very exciting)	Mode = 4 (neither boring nor exciting)
Did you find the tasks… (very difficult – easy)	Mode = 4 (neither difficult nor easy)

Table 15. Results of QIs for Tasks.

In terms of mental effort, the students reported using significantly less effort (m=5) than the Slovene professionals (m = 6.5).

Given the small group sizes, these results are unreliable and hard to interpret. However, they indicate that the tasks are suitably mainstream to suit both academic and professional settings as they were found to be neither too difficult nor too easy, and respondents were able to follow the instructions "neither easily nor with difficulty". At 5.6, the mean amount of mental effort seems to be a little on the low side, although the literature provides no definitive information as to what level constitutes germane load. The students found the tasks less demanding than the professionals possibly because they were used to being taught in EN and taking part in this type of activity.

- Materials in general

In addition to the evaluation process for individual types of training material, overall evaluations were also sought for the materials in general. These were evaluated at ME5 and field tested at various courses and are reported below.

Materials in general (formative): OPEN: UA/UV

A two-day workshop on AD for live events was held by OPEN which is the Expertise Centre for Accessible Media & Culture at the University of Antwerp, 6th – 7th Nov 2018. 6 people attended, all working in theatre. The workshop was taught by Aline Remael, Nina Reviers and Louise Fryer from two of the partner organisations of ADLAB PRO (UA and UV), with input from Hanne Roofthoof, a PhD student at UA. The languages of the workshop were EN and NL. As the workshop was held before all the training material was complete the resulting evaluations should be considered formative.
The 6 students completed an evaluation form.

Summary of Results

This was a small group in line with the group sizes found in IO1's survey of AD courses taught for live AD. Although participants were not specifically asked to evaluate the training materials, in general they were

happy with the content of the course most commonly finding it interesting and well-structured, and that it increased their confidence in AD. All strongly disagreed that it was confusing and all strongly agreed that they would like to know more about AD. Their qualitative comments illustrate the impact of other aspects on learning such as learning environment, including peer group, and the classroom (Fraser, Aldridge, & Adolphe, 2010). Positive comments included "It was briefly framed theoretically and I liked that there was a practical link." "A super team of driven teachers; a lot of time for discussion and participants' input." "Nice to learn about a discipline other than film." "Even if it was very practical it was very useful for me although I'm not a describer" "nice assignments that give a good impression of the challenges and difficulties" "the balance between theory and practice was ideal. Also the option to ask questions at all times, building on the experience of experts". Some of the negative comments concerned technical difficulties that were outside the trainers' control. However practical suggestions included "Maybe more examples of theatre performances with AD to get a better image" and "For a few of the exercises there was not enough time."

Responding to the Evaluation

The negative comments prompted the following actions for Module 3: An Additional Video was created adding AD to the examples of live events so future students could be shown the examples with and without AD, Longer time estimates were included for relevant Tasks.

Materials in general ME5: Session 1

Having been shown and asked to evaluate a reading list, a trainer's guide and an introductory video to the whole course, the 32 participants at Session 1 completed a final evaluation. This final feedback was qualitative and elicited in the form of open questions (any other comments).

Summary of results

Replies were very enthusiastic about the IO4 materials and about the usefulness of the event, although some participants found it hard to provide more thorough feedback based on the short presentation. Positive key words include: excellent; very useful, very educational, flexible, comprehensive, free. The only negative comment concerned the appearance of the animated character in the introductory video who was deemed to be "distracting".

Participants were asked whether they would recommend the training materials and to give a reason for their choice. The majority (97%) gave a positive reply. The only participant who did not tick yes replied n/a to this question. This might be because they were not in a position to recommend materials.

Reasons given for such positive feedback for the materials in general were similar to those given for specific types stressing the clarity, structuring and comprehensiveness of the content, as well as the sound methodology that has led to the creation of open-access free materials.

Materials in general ME5: Session 2

Having been shown and asked to evaluate core and additional videos, and a task, the 36 participants in Session 2 completed a final evaluation of the training materials in general. It was assumed that those attending Session 2 had also attended session 1 and had therefore been exposed to all the types of material. Their final feedback was requested in the form of evaluating content in general on the basis of 5 QIs (interest;

structure; confusion; ease of understanding and increasing understanding of AD) using a 5 point Likert scale. In addition there was an open question (any other comments) and participants were asked if they would recommend the training materials (Yes/No) and to give a reason for their answer.

Summary of results

The quantitative data report means above 4.5 on a 5 point scale for the positive indicators suggesting that participants agreed or strongly agreed the materials were interesting (4.8), well-structured (4.6) and easy to understand (4.7). Most strongly agreed the materials increased their understanding of AD (4.5) and most disagreed the materials were confusing (1.19).

Regarding qualitative feedback, participants were very satisfied with the amount of materials, their quality and their accessibility and were looking forward to having access to them, examining them in more detail and using them. Less positive aspects referred to the varying quality of the videos and the fact that the English could be improved. 16 participants (44%) failed to provide an answer, leaving the "liked least" section blank. In response to whether they would or would not recommend the materials, 15 (41%) participants failed to respond; of those who replied, 20 said that they would and one said they would not.

Discussion

The "shop window" style of evaluation offered by ME5 was not ideal as participants made their evaluations based on a brief introduction to and showing of the materials. This might be reflected in the high number of non-responses to the question as to whether or not participants would recommend the materials. However, the qualitative responses from both sessions were overwhelmingly positive.

Materials in general: Interviews (ME5)

Results

In answer to two binary questions, the trainers interviewed after ME5 unanimously stated (9/9) that they would use the materials in a future training course and that they would recommend them to another trainer (9/9). This positive response is supported by their comments. The comments are divided broadly into strengths and weaknesses in line with the questioning.

Strengths of the Training Materials

Interviewees were enthusiastic about using the training materials either to augment the trainer's own materials or materials currently available in their country. They were also deemed to have the potential to support self-learning:

> **Interviewee 01** "when we do the courses, of course I have some videos I show to them and if I can say in addition you'll find more information here on the website online because courses normally are three to four days. So there is some space in between so that people have questions so I can tell them. OK. There's another source to get information."

> **Interviewee 03** "I have my own materials and so but I think there are some extra things that would be very nice to use or work on their own for students."

Interviewee 03 "I have only seen three tasks or three videos of each but I'm really looking forward to seeing the rest because I'm sure there will be very interesting things and they can be like extra things you can using your lessons."

There was also a perception that isolated trainers would use the materials to support their own learning or practice by seeing how AD is taught elsewhere:

Interviewee 02 "The opportunity to see how other people approach it, how other people frame it in terms of competences is really useful."

Interviewee 05 "So in Australia there is no formal university training in any aspect of audio description at all. So yes it would be brilliant."

Interviewee 05 "I actually want to put myself through this entire course."

Interviewee 09 "Because I'm working in Hong Kong all materials that I can get access to are Chinese materials. So I do want some references from outside the Chinese community."

In particular, some of the additional videos could enable the voices of AD users to be heard in training situations where it was not possible to involve both sighted and blind or partially sighted trainers.

Interviewee 03 "You know it's interviews with users and I miss that in my lessons because you cannot have a blind person of course. And it's always difficult to have this point of view from users and this is already recorded. "

There was also an appreciation of the decision to make the materials freely available under a creative commons licence:

Interviewee 04 "Because of the nature of where we're at, at least speaking for Canada. It is so useful to have something that is not proprietary that can be disseminated freely that has what I would call it I guess a scientific background behind it. We are existing in a vacuum of information and it has bred some unhealthy practices."

However, appreciation of the training materials was not limited to those working where AD is relatively new.

Interviewee 06 "I find them hands on varied creative and based on experience both teaching experience professional experience and connections with the people with the users."

Interviewee 07 "This is good stuff for anybody anywhere."

The training materials were specifically created to be flexible and modular, and this aspect was also received positively:

Interviewee 08 "I think that they would be very useful especially knowing that it's a framework. And not a rigid set of things that it has to be this way but I think it's great to have it organized in a way that people can then take off from."

The materials were especially valued because of the perceived calibre of the ADLAB PRO team:

> **Interviewee 02** " I think having that source material that is that is being held in an institutional context and is being contributed to by a global team of experts is going to make it a lot easier and hopefully get people from zero to describer a little faster and a little more smoothly."

> **Interviewee 07** "I know that Kathy Zeiger who's .. going to be there in charge of the [American] certification effort because they certify others and she wants to have some sort of association with she says, a university. Well this is multiple universities! "

The training materials have been created in English (EN), with subtitles in partner languages where possible. Advice in the trainer's guides urges trainers to localise content as far as possible. It was encouraging to see that many of the trainers interviewed were already considering this. Localisation was commented on by over half of the interviewees (5/9). This was not limited to non-native speakers. Native EN speakers were also considering personalizing the content or tailoring it to their own training circumstances.

> **Interviewee 01** Of course I have to think about how I can use them in English or maybe I can do a voiceover in German so that I adapt this to make it make it more suitable for me.

> **Interviewee 02** There'll probably be some work to do to localize the content.

> **Interviewee 09** First of all maybe I need to translate the materials into Chinese [] and I also did reception studies so I will also input the findings that I have got from my research so that would be a lot of effort because what I can see that there's like a load of information from ADLAB PRO and I don't think I can really translate all the videos into a Chinese context but in Hong Kong one of the advantage that we have is well, many people are bilingual so they can understand English materials. So probably I don't have to spend too much effort on translating anything in English but then I can give some input of my personal experience or my audio description experience that I have, I mean on the Hong Kong market to my students.

Weaknesses

As outlined above, interviewees were actively questioned about perceived weaknesses in the materials. Their responses fell into two main categories, technical presentation and comprehensibility.

Technical Presentation

> **Interviewee 01** "Well, I as a broadcaster was a bit disappointed with the sound quality of the audio. This is… I'm used to that we produce this in the studio and of course this was all done by smartphones or whatever not in really good surroundings so I would do something about that [so] that it sounds okay."

> **Interviewee 06** "I would like to have a look at them more carefully because I didn't have time to but maybe in some of the videos some images could be included. Since these videos are meant for sighted trainers so I think that in those cases images can help illustrate them but fine, I mean I really enjoyed the session. I really enjoyed looking at the materials."

Comprehensibility

> **Interviewee 05** "One of the comments I put on my form was to ask if the training videos themselves could be re-voiced in... It's a rather parochial comment but I think that it's quite important that perhaps these be re-voiced by a native speaker of the even in English like I so am very very grateful that these resources have been done in English because I'm a complete philistine and have no other languages other than English. So it means they're completely accessible to me which is fabulous. But listening to just a couple of them different voicings by different Spanish speakers speaking English it takes a while to cue into their accent and I just wonder if it would be more helpful for an Australian student of audio description to have those main content videos re-voiced."

> **Interviewee 05** "It did include some jargon some inaccessible language and I, I made the suggestion that it all be proof read and also looked at for some of this technical language."

> **Interviewee 07** "It seemed most suited for a university or academic kind of setting and I'm not coming from that place really."

However, the same interviewee also expressed the opinion that the content would be most suitable for students on being first introduced to AD.

> **Interviewee 07** But the, the material, the essence, the content would be fine. I think probably for level one beginning folks principally at least from what I saw today.

Suggestions

Interviewees were also asked to expand on ways in which the materials could be improved.

> **Interviewee 06** "Some adaptations for online teaching would be… would be good. Well in the tasks of using description or multiple choice if they were devised as moodle or moodle- like questionnaires or survey monkey or similar these would be more entertaining. And ready to use. And also not so easy to see the solution. I would use them mainly as reinforcing exercises but it's good that people get immediate reward from having the result just after they do [it]

> **Interviewee 08** Well from my point of view of course I'd always want there to be some aspect about movement.

> **Interviewee 08** received immediate reassurance from the interviewer that the IO4 training materials contain two videos relating to movement (one Core Video and one Additional Video relating to AD of dance, both in Module 3). However, the comment has been included here to illustrate the limitation of evaluating only a sample of the materials.

Responding to the evaluation

The other suggestions and comments relating to perceived weaknesses were addressed at the Trans Project meeting (TPM) held by Partners the following day. It was recognised that these interviews represent summative evaluations which take place once a project (in this case IO4) has been completed and does not allow for remedial action. As detailed above all IO4 materials underwent detailed formative evaluation at the point of creation. For that reason and for the practical considerations that ADLAB PRO is only a few months away from ending and the budget for IO4 has been spent, it was decided that it would not be feasible to re-record the voice-overs at this stage, nor to include more images in the ppt. presentations. Furthermore, the

customisable nature of the materials as emphasised above meant this could be done by individual trainers where there were genuine concerns. Similarly, there is nothing to prevent a trainer from re-creating the tasks online where training is not, or not wholly, conducted face-to-face.

Materials in general: (student/professional evaluations) UAM/RTV-SLO

In addition to evaluating specific training materials (e.g. Core Videos, Additional Videos and Tasks), the students at UAM and the describers at RTV-SLO also completed evaluations for materials in general. They evaluated 4 QIs (interest, curiosity, difficulty and comprehension) on a 7 point labelled Likert scale as well as mental effort on a 9-point scale.

Summary of Results

As explained above, the obvious limitation of this analysis is that two very different, small cohorts have been joined together who were evaluating different stimuli. Yet independent sample t-tests showed their responses to be significantly different for only one QI, namely the difficulty of the materials. This may be because the measures were not sufficiently sensitive.

Despite these limitations, the results suggest the materials are generally considered to be fairly exciting, easy to comprehend, required a moderate amount of mental effort and that participants were very curious to see the rest of the materials. This is consistent with the results found from the other evaluation methods (the "shop window" evaluation; the focus group; the semi-structured interviews and the evaluation questionnaire used at ME5).

FINAL RESULTS

In this section, the results from all the evaluations are considered, giving a comprehensive picture of how the different types of training material were received. These are addressed as strengths and weaknesses before limitations are considered and some final conclusions drawn.

STRENGTHS OF THE TRAINING MATERIALS

Core Videos

Both formative and summative evaluations ensured that the videos are clear, succinct, and easy to understand. The general information they contain is quite easy to remember. They are not confusing and make a moderate demand on most learners, being neither too easy nor too difficult for beginners to follow.

Additional Videos

Of all the materials, the Additional Videos were the most varied and were rated the most highly. Compared with the core videos they were thought to be more interesting and emotionally engaging. Amongst the 112 Trieste students who answered this question (99.1%) would recommend the Additional Video, compared with 107 (94%) who would recommend the Core Video. The 4 audio describers at RTV-SLO, the 6 Polish students and the 36 participants at ME5 also expressed a preference for the Additional Videos over the Core Videos.

After ME session 1 when asked whether they would recommend the training materials, the majority (97%) gave a positive reply. The only participant who did not tick yes replied n/a to this question. After ME session 2 15 (41%) participants failed to respond; of those who replied, 20 said that they would and one said they would not. This corresponds to 95% of those who replied but only 55% of the total.

Results

The evaluations carried out for IO5 were subject to a number of limitations. There were inconsistencies in the EFs (e.g. some used unlabelled 1- 5 Likert scales and some asked for Yes/No responses and used a 1- 10 scale others were labelled and scaled 1 – 7). While this was useful for demonstrating types of evaluation method that fed into The Guide, in terms of the project, it made it difficult to compare evaluations. In addition, with the exception of the student evaluations in Trieste, response sets were generally too small for quantitative evaluation to be appropriate. As these small sample sizes were expected, the mixed methods approach compensated by offering qualitative data. The focus group was also small (5 participants) but within the range thought acceptable by Brown (1999). Even the limited amount of qualitative data introduced by soliciting what participants liked best and least about the materials provided a much more concrete indication of response, than simply answering Yes/No to the question of recommendation, or even the means and modes from the QIs described above.

On the positive side, compared with the limited amount of qualitative data generated by the EFs, the wealth generated by the semi-structured interviews and the focus group shows how much fuller the information can be when the opportunity arises for follow-up questions leading to further elaboration. For example, the use of Core and Additional Videos by PSL as tools for advocacy and training had not been foreseen and would not have been uncovered by a purely quantitative evaluation.

GENERAL DISCUSSION

This chapter has demonstrated the various contributions of a number of evaluation strategies employed for IO5. They were designed not only to provide a rich source of cumulative data but also to show the advantages and limitations of various evaluation approaches. These are discussed in The Guide. This report draws some final conclusions about the quality of the IO4 materials produced by ADLAB PRO.

As explained in Chapter 1, IO5 took a mixed methods approach to evaluation. The quantitative results were consistently good, with means above 4 for all material types from all evaluations. However, as expected some response sets were too small for quantitative evaluation to be appropriate. There were also limitations.

CONCLUSION

The purpose of the IO5 was to guarantee the quality of the training materials created, to ascertain their effectiveness, usability, and consequent longevity. Types of evaluation were mixed and wide-ranging. Perhaps the most surprising feature of the multiple types of evaluation used in this project has been the consistency of the responses, be they from AD students in Poland, AVT students in Italy, professional describers in Slovenia or the trainers, researchers, academics and providers of AD content from across the globe who attended ME5. By combining them, we can be confident that the ADLAB PRO training materials are well-structured, easy to understand and not confusing. AD users feel that PSL are appropriately represented and that the videos will be useful for their own training purposes. We also know that the videos are not of the highest quality in terms of production values and that trainers will have to localise some of the materials according to their own needs. It is for that reason that they have specifically been designed to be flexible. We also know that they fulfil a need and that there is an appetite for them amongst trainers. As one interviewee expressed it: “This is good stuff for anybody anywhere.”

REFERENCES

ADLAB PRO (2017a). Report on IO1: Assessment of Current AD Training Practices. Retrieved from http://www.adlabproject.eu

ADLAB PRO (2017b). Report on IO2: Audio description professional: Profile definition. Retrieved from http://www.adlabproject.eu

ADLAB PRO (2018). Report on IO3: Course Design. Retrieved from http://www.adlabproject.eu

ADLAB PRO (2019). Report on IO4: Development of course content. Retrieved from www.adlabproject.eu

Campbell, R., Quilty, B., & Dieppe, P. (2003). Discrepancies between patients' assessments of outcome: qualitative study nested within a randomised controlled trial. British Medical Journal, 326, 252-253

Chemlimsky, E. (1997). The coming transformation in evaluation. In E. Chelimsky & W. Shadish (Eds.), Evaluation for the 21st century. Thousand Oaks, CA: Sage

Clarkson, M. (1994, May). A risk based model of stakeholder theory. In Proceedings of the second Toronto conference on stakeholder theory (pp. 18-19)

Collins, K. M., Onwuegbuzie, A. J., & Jiao, Q. G. (2006). Prevalence of mixed-methods sampling designs in social science research. Evaluation & Research in Education,19(2), 83-101

De Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. Instructional Science, 38(2), 105-134

Frechtling, J. (2002). The 2002 User-Friendly Handbook for Project Evaluation

Fryer, L. & Cavallo, A. (2018) "Integrated Access Inquiry 2017-18 Report." retrieved from extant.org.uk/docs/uploads/Extant_Integrated_Access_Report_2018_Full_Length.docx

Fryer, L., & Freeman, J. (2014). Can you feel what I'm saying? The impact of verbal information on emotion elicitation and presence in people with a visual impairment. In Challenging presence: Proceedings of the 15th international conference on presence (pp. 99-107)

Goyder, J. (1985). Face-to-face interviews and mailed questionnaires: The net difference in response rate. Public Opinion Quarterly, 49(2), 234-252

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. American Sociological Review, 447-462

Kennedy, D. (2007). Writing and using learning outcomes: a practical guide. Cork: University College Cork

Kirkpatrick, D. L. (1959). Techniques for evaluating training programs. Journal of ASTD, 11, 1–13

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. Presence: Teleoperators & Virtual Environments, 10(3), pp. 282-297

Mahapatra, R., & Lai, V. S. (2005). Evaluating end-user training programs. Communications of the ACM, 48(1), 66-70

Marsden, J. (1991). "Evaluation: Towards a definition and statement of purpose." Australasian Journal of Educational Technology 7.1

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Academy of Management Review, 22(4), 853-886

O'Cathain, A. (2009). Mixed methods research in the health sciences: A quiet revolution

Pinsonneault, A., & Kraemer, K. (1993). Survey research methodology in management information systems: an assessment. Journal of management information systems, 10(2), 75-105

Scheirer, M. A. and Schwandt, T. (2012). Planning evaluation through the program life cycle. American Journal of Evaluation, 33(2), 263-294

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. Educational Psychology Review, 10, 251–296

Wigley, J. (1988). Evaluating training: Critical issues. Training and Development in Australia 15(3), 21-24

Wilkes, M., & Bligh, J. (1999). Evaluating educational interventions. Bmj, 318(7193), 1269-1272

Windsor, D. 1992. Stakeholder management in multinational enterprises. In S. N. Brenner & S. A. Waddock (Eds.), Proceedings of the Third Annual Meeting of the International Association for Business and Society: 121-128. Leuven, Belgium

DISCLAIMER

Publication and preparation of this report was supported by ADLAB PRO (Audio Description: A Laboratory for the Development of a New Professional Profile), financed by the European Union under the Erasmus+ Programme, Key Action 2 – Strategic Partnerships, Project number:2016-1-IT02-KA203-024311.

The information and views set out in this report are those of the author and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

ACKOWLEDGEMENTS

This report was written by UV. Thanks to those who took part in the Focus Group at RNIB and to the students and learners who took part in field testing.Thanks, too, to all our evaluators and interviewees: Dr. Bernd Benecke, Dr. Cristóbal Cabeza-Cáceres, Matthew Cock, Prof. Elena di Giovanni, Dr.Eva Espasa, Prof. Deborah Fels, Esther Geiger, Anne Hornsby, Rachel Hutchinson, Prof. Jan Luis Kruger, Dr. Dawning Leung, Dr. Mariana Lopez, Dr. Eleanor Margolies, Will McRostie, Prof. Vincenza Minutella, Carmen Oven, Dr. Juan Pedro Rica, Mary Plackett, Dr. Anna Sadowska, Kari Seeley, Rebecca Singh, Dr. Kirstin Smith, Dr. Joel Snyder, Kim Starr, Dr. Agnieszka Szarkowska; Dr. Rebecca Tipton, Dr. Maria Valero, Maya Van Puymbroeck, Susanne Verberk, Shak Yousaf.

HOW TO QUOTE THIS DOCUMENT

You can quote this report as follows: ADLAB PRO (2019). Report on IO5: Evaluation and Testing. Retrieved from www.adlabproject.eu