

EVALUATION AND TESTING (IO5)

Louise Fryer
Utopian Voices Ltd.



Barcelona, TPM, 22 March 2019

Dept. of Legal, Language, Translation and Interpreting Studies, Section of in Modern Languages for Interpreters and Translators
University of Trieste, Via Filzi, 14 - 34144 Trieste, Italy
Project numberStudies: 2016-1-IT02-KA203-024311
www.adlabproject.eu

FUNDED BY THE ERASMUS + PROGRAMME OF THE EUROPEAN UNION





IO5 DETAILS

- **Full title:** Evaluation and Testing.
- **Duration** initially envisaged running intermittently M7 – M36
 - In fact: M1 – M36 (Oct 16 – Sept 19)
- **Leading partner:** UV (Fryer)
- **Contributing partners:** All
- **Other contributors**
 - All stakeholders: participants in ADLAB PRO events/AD providers/ AD users /AD students, trainers/lecturers/course deliverers.



EVALUATION FOR ADLAB PRO

- Evaluation ÷ Internal (Management/Project evaluation)
 - External (Output evaluation)



ACCORDING TO THE APPLICATION FORM

- Evaluation should be built in from the start.
- Course content (IO4) will be designed in progressive form (tested at each stage to secure content quality, adequacy and progress.)



COMPLETED SO FAR...

Material Type	Sent to	Date returned	report
Core video	Dr. Cristobal Cabeza-Caceres	30.04.18	180430_UV_IO5_training mats_evaluation_Report_Core_video.doc
	Dr. Agnieszka Szarkowska	23.04.18	
	Dr. Bernd Benecke	16.03.18	
Introductory video	Prof. Maria Valero	21.11.18	231118_UV_IO5_IO4_Materials_evaluation Report_introductory_video.doc
	Ms. Kim Starr	22.11.18	
	Mr.Shak Yousaf	22.11.18	



COMPLETED SO FAR CONT.

Material Type	Sent to	Date returned	report
Sample task M2U5	Prof. Juan Pedro Rica	04.11.18	Final version: 181203_UV_IO5_Evaluation report sample task_final.doc
	Dr. Anna Sadowska	03.11.18	Interim report written & circulated 03.12.18 181203_UV_IO5_IO4_Materials_Evaluation Report_sample task_interim.doc
	Prof. Vicenza Minutella	30.11.18	



SAMPLE TASK RESULTS (EXTERNAL EVALUATIONS)

- All the external evaluators agreed that the tasks are well structured and that their students would find the tasks interesting.
- The demands of the task appear appropriate in terms of cognitive load, which at $m=6$ is neither too easy nor too demanding (de Jong, 2009).
- All the external evaluators showed an appetite for more tasks like these from ADLAB PRO.



QUESTIONS ARISING

- Should the tasks only be based around the core videos?
- Should the number of questions be limited to 5?
- Should the multiple choice questions have a uniform format - with just one correct answer?
- Should the key be off limits to students (e.g. contained in the trainer's guide)?
- Is a key necessary for more open-ended tasks or do we need example solutions?
- Should the time estimate apply only to the whole task (not its constituent parts)?
- Should the task description be less detailed ?



CONCLUSIONS AFTER UAB PETITIONED THE ACADEMIC PARTNERS BY EMAIL

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



EVALUATING STUDENT PROGRESS FOR PEDAGOGICAL PURPOSES.

- UV has produced and been testing EFs for assessing student performance on 4 measures of AD Quality (delivery; accuracy; language; synchrony) adapted from interpreter rating scales. Can be used for self-; peer-based assessment or teacher assessment.
- A chapter explaining the rationale: Fryer (2019) Quality Assessment in Audio description: Lessons learned from Interpreting. In E. Huertas-Barros, S. Vandepitte and E. Iglesias-Fernández (Eds.) Quality Assurance and Assessment Practices in Translation and Interpreting. IGI-Global.



A SAMPLE EVALUATION TASK FOR M3_U7

- Created Evaluation sheets/rubric – commented on by partners.
- Exchange your AD script with a colleague. Using the ADLAB PRO evaluation sheets, evaluate each other's work.
- 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.
- Piloted with 9 MA students @UCL 30.10.18.



RESULTS

- The task has improved my understanding of the needs of AD users: **m = 4.22**
- I found the task interesting: **m = 4.22**
- The task has increased my understanding of audio description: **m = 4.44**
- The demands of the task were appropriate: **m = 4.67.**
- Where 1= minimal effort and 9 = extreme effort, how much mental effort did it take to complete the task? **m = 7.11**
(range = 6 – 8; **mode = 7**)



MENTAL EFFORT BY MOTHER TONGUE

sig. dif (chi-square, $p = .048$)

Effort score		6	7	8	Total
	EN	0	3	1	4
	European	0	0	2	2
	Chinese	2	1	0	3



QUALITATIVE COMMENTS

- 'Where I would concentrate a lot on describing outfits, room layouts and expressions, my peers would focus strongly on just actions. That made me think that maybe describing too much too fast may result in too much information for the viewer/listener.'
- 'I changed the phrase "suck it up" into the word "snort", which sounds more natural and appropriate in this context.'
- 'I thought the peer evaluation was an excellent and very useful task because it made me realise the many ways in which others perceive film and actions.'



TRACING THE EVOLUTION THROUGH THE IOs

- The ability to give and receive criticism is one of the soft skills highlighted in IO**1**.
- An example of team-working skills - considered important or extremely important by 60% of existing describers and 73% of service providers in IO**2**.
- Included as a competence and a unit in IO**3**.
- Task relating to evaluation created for IO**4**, having been tested by IO**5**.
- Feedback through IO5 affects task timing estimates.
- This will aid the certification process (ECTS/ECVETS) in IO6.



WHOLE COURSE EVALUATION (MOSTLY M3; M1)

- 5 day course delivered to 6 learners in Cardiff (organised with Taking Flight Theatre Company 1-5th October 2018).
- Materials used: Powerpoints, core videos and materials later incorporated as Additional Videos.



LIMITATIONS

- Not all materials ready (too early).
- But you can't teach only part of a course.
- Ecologically valid.
- Materials designed to supplement existing materials as well as provide a complete course.
- Small scale (only 6 participants) in line with the non-academic courses sampled by IO1, which found that "Almost half of these courses is addressed to small groups of fewer than 10" (Chmiel & Mazur, p.10).



Statement	score out of 5 (mode)
The training has changed my approach to AD	4
The training was well structured	4
I found the training interesting	5
I found the training confusing	1
I would like to know more about AD	5
The training has increased my confidence as an audio describer	5

RESULTS



TWO MORE QUESTIONS

- The academic level of the content was too difficult/too simple/ **about right (unanimous)**.
- The amount of content included in the time available was too much/too little/**about right(unanimous)**.



IN THE APPLICATION FORM

- Also **focus groups** and with service users will take the form of quality assessment of audio descriptions prepared by course participants at the end of the course.



MOST PRECIOUS FEEDBACK FROM A PSL

- 'The show I'd say is quite interesting. It moved very quickly within time and I think the beginning, for me, when Mary comes into the lab and they meet for the first time. For me emotionally I think it was a bit weak especially on Chris's part. It just seemed like they just got too comfortable with each other, too quickly. I liked Chris's and John's relationship, it seemed quite real and.. I don't know what else to say about it.'
- Passes the "bar" test.



WHAT DID YOU FEEL WHEN MARY WENT TO SETTLE THE BABY?

- CC: 'Oh I thought she was going to kill it.'
- LF: 'Good – that is the correct answer!'
- CC: 'It was quite interesting the men seemed to be taking it very, very calmly.'
- LF: 'And what on earth were they doing letting her?'
- CC: 'I know! If I was a parent there's no way in hell, I'd let her go in that room on her own.'
- LF: 'Sadly the describers have no control over that.'
- CC: 'I know but it's great isn't it? I was thinking "Why would you do that?'



QUALITATIVE COMMENTS FROM LEARNERS: 'WHAT I LIKED BEST'

- 'Chance to practice Ading, Ad-ing each other, clips & movement. I also love the energy of the teachers.'
- 'Having a pro-describer and a theatre practitioner AD user makes a knowledgeable lead team. I really enjoyed learning about how to balance AD with dialogue as I've not worked in that way before.'
- 'The atmosphere of support, a fantastic group of people to learn with. Great constructive conversation and content of the course.'



'WHAT I LIKED LEAST'

- Things felt a little crammed together, but I think that was more to do with the fact we had to compress the content into a week, but that was the only thing.
- Echoes in the space and tech issues. I think because some people had to miss the last day it meant feedback e.g. for the TT had to be rushed. This comes with such intensive training as tiredness & vulnerability of performers.
- Not having a day being introduced to all types of AD. I would have loved to learn more if there was less script to prep.



THE FLY IN THE OINTMENT...

- 'The success of the curriculum cannot be determined by the tasks alone. It is a complex interaction between **student aptitude, teacher effectiveness** and **learning environment.**' (Fraser, 1981)



ONE WAY TO COMPENSATE

- To test the same content with other teachers/learners/different venue.



COURSE MATERIALS TESTING (2)

- Industry learners in Antwerp 6th/7th Nov 2018.
- Teachers: LF; AR; NR; HR.
- 6 participants, all working in theatre.
- Materials used: Powerpoints/core videos (from M1 & M3) and materials later incorporated as Additional videos & tasks.
- Aspects covered:
 - Semiotics of live AD.
 - AI.
 - AD scripting and improvisation.
 - Delivery.



RESULTS

Statement	score (mode)
The training was well structured	5
I found the training interesting	5
I found the training confusing	1
I would like to know more about AD	5
The training has increased my confidence as an audio describer	5



QUALITATIVE COMMENTS FROM LEARNERS: 'WHAT I LIKED BEST'

- '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.'
- 'Getting taught by experienced AD persons.'
- 'The balance between theory and practice was ideal. Also the option to ask questions at all times, building on the experience of experts.'



'WHAT I LIKED LEAST'

- 'That the workshop ended :) Thank you!'
- 'Maybe more examples of theatre performances with AD to get a better image.'
- 'The technical preparation: slow equipment now and then.'
- 'For a few of the exercises there was not enough time.'



FEEDBACK INCORPORATED INTO IO4 TASKS AND ADDITIONAL VIDEOS.

- Maybe more examples of theatre performances with AD to get a better image.
- Created AVM3_U1 in response.
- For a few of the exercises, there was not enough time.
- Things felt a little crammed together.
- Factored into estimated Task timings.



MATERIAL EVALUATIONS BY INTERESTED TRAINERS

- ADLAB PRO approached by:
- **Kat Germain** (Audio Description, Consultation & Described Video Specialist, Actor, Film & Theatre Creator, Activist. Location Toronto, Ontario, Canada Industry Performing Arts)
- **Dawning Leung** (PhD student at UCL) founder of AuDeHK.
- **Meg Wedding** (Learning and Development professional. Skilled in Operations Management, Instructional Design, Desktop Publishing, Digital Content Development and Accessibility. Melbourne, Australia.)
- **Jane Brambley** ADA Trainer (UK)



RESULTS PENDING

- Only Kat Germain & Jane have completed an evaluation.



QUALITATIVE COMMENTS

- 'I am so utterly grateful for these materials, even though it might sound like I wasn't by this evaluation! The generosity shown by the AD community world-wide is without comparison. I so appreciated not having to start from 0'. (KG) Recommend? 'Yes'.
- 'These materials seem to me to be audio notes, of more use when structuring a course than teaching it. There is no thought given to the presentation, which is very repetitive – it would not engage me as a student.' (JB) Recommend? 'No.'



BACK TO COMPLETED SO FAR...

Material Type	Sent to	Date returned	report
Reading lists	M1 Dr. Eleanor Margolies	29.11.18	181207_UV_I05_I04_Materials_Evaluation Report _reading lists_final_updated.doc
	M2 Prof. Deborah Fels	30.11.18	
	M3Dr. Sarah Weaver (sick) so Prof. Elena di Giovanni	28.11.18	
	M4 Rachel Hutchinson	26.11.18	
	M5 Prof.J-L Kruger	03.12.18	
	M6 Dr. Mariana Lopez	26.11.18	



RESULTS

- The reading list contained appropriate references **m = 4.8**; (1 evaluator agreed; 5 agreed strongly).
- The number of references was sufficient: **m = 5** (all evaluators agreed strongly).
- I feel I have enough knowledge to assess this reading list. **M = 4.4** (3 evaluators agreed; 2 agreed strongly).
- I would use this reference list if I were teaching a relevant **course m = 4.5** (1 evaluators agreed; 4 agreed strongly; 1 neither agreed nor disagreed).



'I WOULD USE THIS REFERENCE LIST IF TEACHING A RELEVANT COURSE' RESPONSES BY MODULE

Module	Response
M1	Agree strongly
M2	neither agree nor disagree
M3	Agree strongly
M4	Agree strongly
M5	Agree strongly
M6	Agree



SUGGESTIONS ACTED ON.

- Most of the references are from a limited set of European sources; there are very few from North American sources. It would be good to have a broader perspective as indicated.
- Depending on the course aims and the time to be devoted to reading, it may be useful to identify an "essential list" of readings within all the items listed. Originally there were 10 references per unit, 5 of which were subsequently highlighted as essential texts.
- Information should be included in the course guide as to how the reading lists are to be used.



LIMITATIONS

- Evaluators only saw RL for individual modules. No overview.
- Deborah Fels included a long list of her own publications (for M2) some of which were in other modules.



BACK TO COMPLETED SO FAR...

Material Type	Sent to	Date returned	report
Trainers guide	Carmen L. Oven (SLO) carmen.loven@gmail.com	23.01.19	040219_UV_I05_ IO4_Materials_Evaluation Report_trainers_guide
	Ms. Mary Plackett maryplackett@gmail.com	02.02.19	
	Ms. Anne Hornsby mindseyedescription@gmail.com	30.01.19	



RESULTS

The trainer's guide...	Response
...was clear and informative.	M= 4.3 (mode = 4)
...was confusing.	M = 1.3 (mode = 1)
I feel I have enough experience in teaching/training to assess this guide.	M= 4.3 (mode = 4)
...would help me decide whether to use the ADLAB PRO training materials if I were teaching a relevant course.	M= 4 (mode = 4)
is fit for purpose	M= 4.3 (mode = 4)



QUALITATIVE COMMENTS

- 'Straightforward and uncomplicated information'.
- 'Assessment criteria to accompany the learning outcomes?'
- 'I find Module 5 of Trainer's guide very systematic and clear. I did not see other 4 modules but with this one I would get a wider perspective of what could be and should be done to create and maintain high level of audiodescription and other tasks concerning that matter.'



OUTCOME

- The sample guide was reformatted in line with RNIB's suggestions as follows: Add a line break between each content line;
- Use list bullets consistently throughout the document instead of bullet points.
- Assessment criteria to accompany the learning outcomes – rejected as impractical to introduce for a range of course situations.



FROM YESTERDAY'S ME...

- **Quantitative Feedback** from Evaluation forms.
- Other forms of **Qualitative feedback**:
- **One-to-one interviews** with 9 stakeholders.
- 9/9 would use the materials 9/9 would recommend them.



SOME NEGATIVE COMMENTS

- Concerns over sound quality and ease of understanding the non-native commentaries.
- Complex language and jargon.
- Need proofreading.
- Perception that they're more aimed at academic settings.



SUGGESTIONS

- Make the tasks suitable for online platforms e.g. Moodle
- More pictures but teachers can modify the Powerpoint.
- Re-record the voice-overs to improve the sound and comprehensibility (core videos).



DISCUSS

- How shall we deal with negative feedback?
- Act on it?
- Ignore it?
- The best we could do in the time and with the money available.



ON THE POSITIVE SIDE

- Lots of positive comments.
- Quality of the people contributing to the materials – global team of experts.
- ‘(It will) help get people from zero to describer a little faster and a little more smoothly’.
- ‘...extra things you can use in your lessons’.
- ‘This is good stuff for anybody anywhere’.
- ‘I’m very interested in the materials because working in Hong Kong and only have access to Chinese materials’.



ON THE POSITIVE SIDE CONT.

- 'Courses are normally three to four days, so when people have questions I can direct them to the website'.
- 'I think it's wonderful that they're licensed under Creative Commons because of the nature of where we're at...Speaking for Canada, we're existing in a vacuum of information and it's lead to some unhealthy practices'.
- 'I find them hands-on, varied, creative and based on experience'.
- 'The opportunity to see how other people frame [their teaching material] in terms of competences is really useful'.



STILL TO COME

- Focus group with RNIB April 11th
- Proposal: discuss Additional videos relating to sight loss. 3 in M3 (AVM3_U5_1; AVM3_U5_2; AVM3_U5_3. Others?
- Suggestions?



LIMITATIONS OF IO5 SO FAR

- Small number of evaluators for whole course materials.
- Samples evaluated, might not correspond to whole content.
- Some Modules (M1, 2, 3) evaluated more than others (M4,5, 6).



STRENGTHS SO FAR

- Quality and reach of evaluators for sample IO4 materials.
- 15 evaluators top in the AVT field with a global reach from Europe, Asia and North America.
- Range: from all types of stakeholder including researchers, Professors lecturers and tutors of professional courses, from AD users and AD learners from students in academia and vocational learners.
- Ecologically valid. Feedback from actual courses.



STILL TO COME

- Course materials (M1, M2) tested at AMU & AU (trainers & students - academic)
- Course materials(M4) Translation and interpreting school Altiero Spinelli in Milan, Associazione italiana subvedenti Milano. (trainers & students - academic)
- Course materials (M3) tested at UCL (trainers & students vocational).
- Course materials (M1, M2) to be tested with students in Hong Kong at City University, HK and with vocational learners at ADeHK .
- RTV will carry out assessment of selection of materials – deadline?



COGNITIVE LOAD

- Single question included on CL in the “EF training mats cognitive”.
- Other continuous variables (self- assessed) attention; interest; memory; arousal; speed; easy to follow.
- Compare by region; 1st language EN; age; familiarity with AD.



FINAL OUTPUT OF IO5

- **Title: A Guide to Project Evaluation: ADLAB PRO a case study**
- short reports ultimately combined into a handbook (print-based and online), satisfaction survey forms and measurable satisfaction indexes, **methodological guidelines** for testing cognitive load and other reception variables.
- I've made a start on this as a "how to" guide for future projects. TOC in DB.



THE PURPOSE OF THE HANDBOOK.

- Outline the background to the evaluation process.
- Explain how this has been implemented with respect to ADLAB PRO.
- Of interest to anyone considering using the ADLAB PRO curriculum
- Anyone wishing to undertake evaluation of training materials in the future.



HANDBOOK CONT.

- Length: c.25,000 - 30,000 words.
- Format: application form (print & online).
- Partners to help with formatting/ proof-reading/reviewing.
- Timeline.
- Draft produced by 30th April.
- Comments by 6th May.
- Final by 13th May.
- Ready by Ljubljana (3rd/4th June).



REFERENCES

- Blank, R. (1993) Developing a System of Education Indicators: Selecting, Implementing, and Reporting Indicators. *Educational Evaluation and Policy Analysis*, 15(1, Spring): 65-80
- Foxon, M (1989) Evaluation of Training and Development Programs: A review of the literature. *Australasian Journal of Educational Technology*, 5 (2)
- Frechtling, J. (2002) *The 2002 User-Friendly Handbook for Project Evaluation*
- Fraser, B.J. (1981) Learning Environment in curriculum evaluation: A review. *Evaluation in Education Jan 1*; 5(1): 1-93



REFERENCES CONT.

- Schmidt, H. G., Dolmans, D., Gijselaers, W. H., & Des Marchais, J. E. (1995). Theory-guided design of a rating scale for course evaluation in problem-based curricula. *Teaching and learning in medicine: an international journal*, 7(2), 82-91.