



A Model for Erasmus+ Expert Training

Strengths, Successes and Areas for Future Development: Dual Perspectives

EVALUATION REPORT 2015



A Model for Erasmus+ Expert Training

**Evaluation of Transnational Cooperation Activity
involving Experts from across Europe in Joint
Training and Assessment Activities.**

Report prepared by Paul Guest (Consultant) in
conjunction with Erasmus+ National Agencies in
Iceland, Norway and Sweden.

December 2015



Contents

1.	Preface	4
2.	Background and History	5
3.	Evaluation Methodology	5
4.	What the People Say	7
5.	What the Data Says	13
6.	Conclusions	31
	Appendices	32

1. Preface

This evaluation report has been produced as a part of the Transnational Cooperation Activity entitled “A Model for Expert Training”, an initiative led by Erasmus+ National Agencies in Iceland, Norway and Sweden.

The report attempts to bring together two different evaluation perspectives: on the one hand, considering “what the people say” through reviewing quantitative and qualitative feedback, gathered from experts participating in expert training in 2015; on the other hand, reflecting on “what the data says” through examining and interpreting data tied to joint expert assessment activity in three European countries (Iceland, Norway and Sweden). In all cases, the goal is simple and centres on continuous improvement, highlighting the successes of the past and confirming goals and ambitions for future development and delivery of expert training within the Erasmus+ programme.

Report content, findings and conclusions were initially produced with a view to prompting discussions during a final review meeting (Stockholm, July 2015), involving senior staff from each of the three core National Agencies and a seasoned expert from Scotland.

2. Background and History

The original “Model for Expert Training” goes back to the start of the Erasmus+ Programme at which point National Agencies (hereafter NAs) were invited to propose one or more Transnational Cooperation Activities (hereafter TCAs) aimed at strengthening the delivery of the Erasmus+ programme across Europe. As an integral part of Erasmus+ programme delivery, TCAs complement and provide added-value to Erasmus+ programme fields and actions and target broader, systemic impact that is aligned with the goals, objectives and priorities of Europe 2020¹, ET2020², the EU Youth Strategy³, and the annual European Commission Work Programme⁴.

Having already developed and delivered the “Model for Expert Training” in 2014, the *SOAR* model of evaluation was subsequently adopted with a view to reflecting on the *Strengths* (of the model), *Opportunities* (brought forth through transnational collaboration), *Aspirations* (for future development and delivery) and *Recommendations* (for change and improvement). Feedback was positive, constructive and future-focused and there was clearly no shortage of ambition for continued collaboration.

In the autumn of 2014, activities centred on continued promotion of the successes of developing and delivering the “Model for Expert Training” within distinct training events - events targeted at experts involved in the assessment of Key Action 1 (hereafter KA1) and Key Action 2 (hereafter KA2) applications - and on confirming commitments from core NA partners for delivery of an extended “Model for Expert Training” in 2015. Notably, interest was already being shown, at that point, beyond the original core group of NAs.

By early 2015, revised goals and objectives had been agreed for the TCA, including:

- o to further develop the “Model for Expert Training” for (internal⁵ and external) experts involved in the assessment of KA1 and KA2 applications;
- o to host two transnational training sessions (for KA1 and KA2 respectively) for selected experts from NAs across Europe, with participants subsequently expected to be involved in the training of other experts involved in the evaluation of KA1 and KA2 projects in their respective countries;
- o to continue to promote the “Model for Expert Training” to other Erasmus+ NAs, the European Commission and those working on other EU programmes and actions not having participated in 2014⁶;

Related outputs and outcomes were confirmed as:

- o a refined model for the training of experts, including updated materials, exercises and slides;
- o common transnational training sessions for KA1 and KA2 experts.

In February 2015, a planning meeting was held in Stockholm, involving staff from the three core NAs and an expert consultant from Scotland⁷, confirming plans for the delivery of a series of training sessions (hereafter, Expert Training Sessions) in 2015, alongside areas for continuing development and improvement of the training materials and training delivery.

In March 2015, a KA1 Expert Training Session took place, in Reykjavik, and involved 46 participants from 16 countries. In April 2015, a KA2 Expert Training and Assessment Session took place, in Stockholm, and involved 116 participants from 20 countries. In both instances, an online survey was subsequently launched, inviting (internal and external) expert participants to provide quantitative and qualitative feedback, with results considered within the current evaluation report (c.f. Chapter 4: What the People Say).

Having delivered these events, assessment data was subsequently gathered from the *Online Expert Evaluation Tool* (OEET) with a view to reflecting on change or improvement that may or may not be attributable to the “Model for Expert Training” (c.f. Chapter 5: What the Data Says).

1 http://ec.europa.eu/europe2020/index_en.htm

2 http://europa.eu/legislation_summaries/education_training_youth/general_framework/ef0016_en.htm

3 http://ec.europa.eu/youth/policy/youth_strategy/index_en.htm

4 http://ec.europa.eu/atwork/pdf/cwp_2015_en.pdf (Commission Work Programme 2015)

5 in many cases, Erasmus+ NA staff also operate as “internal” assessors thus event participation is also extended to these staff

6 including Erasmus+ actors working in the field of HE, where there had been a previous decision not to participate

7 the same Consultant that had provided support in developing and delivering the previous “Model for Expert Training” in 2014

3. Evaluation Methodology

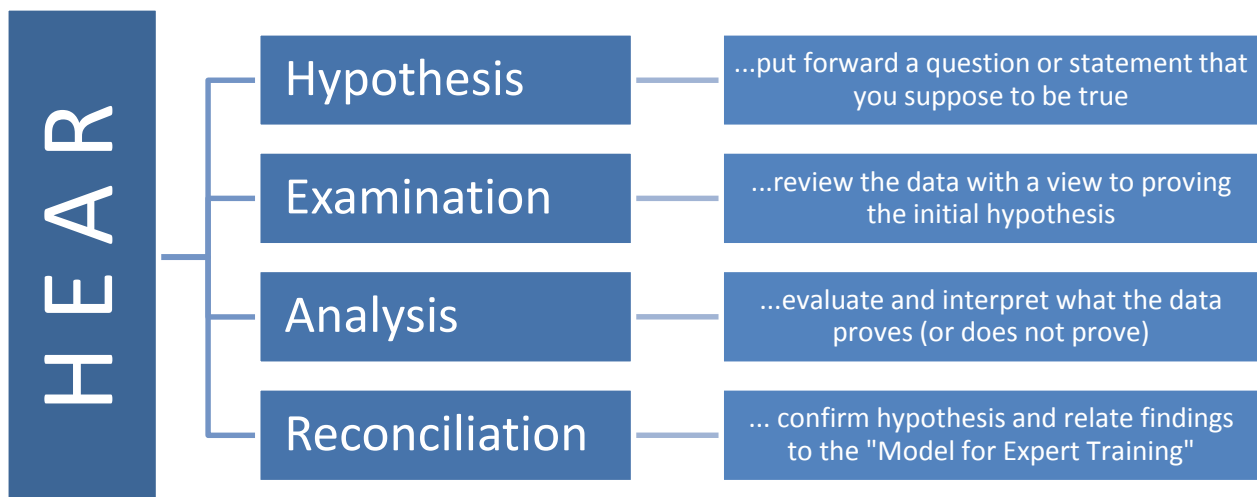
With evaluation having relied heavily in 2014 on thoughts and perspectives brought forth by participating experts - through onsite feedback, an online survey and a series of telephone interviews - it was agreed to adopt an alternative evaluation methodology in 2015, centred on a two-tiered approach to data collection and analysis.

On the one hand, experts (internal and external) would be given a chance to reflect and provide feedback on the perceived benefits and usefulness of their participation (c.f. Chapter 4: What the People Say).

On the other hand, a more analytical approach would be adopted in which data from 2014 and 2015 assessments (individual and consolidated) would be compiled and considered with a view to confirming change and improvement that is potentially attributable to the "Model for Expert Training" TCA (c.f. Chapter 5: What the Data Says).

Input from participating experts - considered during the current evaluation - relied on *QuestionPro*, an online survey tool, with tailored question sets developed by core NAs and circulated to those participating in the KA1 Training Session (Reykjavik, March 2015) and the KA2 Training and Assessment Session (Stockholm, April 2015).

Additional assessment-related data was sourced by the core NAs in Iceland Norway and Sweden, and extended to KA1 and KA2 assessment activity having taken place in both 2014 and 2015. Mindful of the lack of comprehensive baseline data and of the relative novelty of programme data - covering only the first two years of the new Erasmus+ funding programme - evaluation activity centred on the HEAR approach (below) through which the gathered assessment-related data could be considered, presented and subsequently interpreted.



Initial findings were presented to a meeting involving representatives from the three core NAs (Stockholm, July 2015), with a view to verifying initial hypotheses and discussing and agreeing on variables that would need to be additionally referenced under analysis and/or reconciliation.

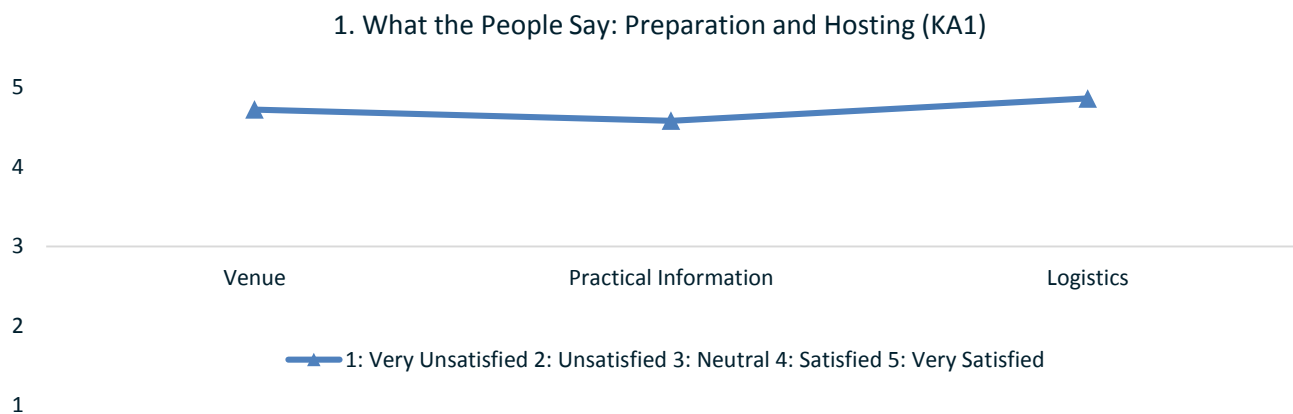
4. What the People Say

In this section of the evaluation report, we consider feedback that was gathered following each of the two Expert Training Sessions held in Reykjavik (March 2015) and Stockholm (April 2015). Data was gathered using *QuestionPro*, an online survey tool, with unique, yet similar, question sets developed for use following each of the two expert training sessions.

With the ultimate goal being to provide a rapid insight into perceived successes, confirming all or any areas worthy of enhancement or improvement, minimal evaluative interpretation is given in this section, presenting instead the perspective of survey respondents.

Expert Training Session for KA1

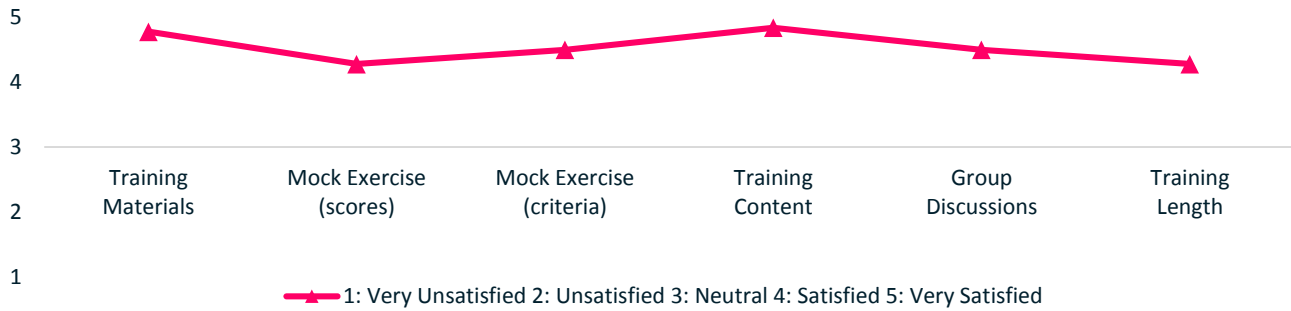
Following the one-day expert training session in Reykjavik (March 2015), held over an evening and a full day, a series of questions were sent to participants with a view to securing input on issues tied to event hosting and event delivery, for which participants were asked to give ratings using a 5-point scale. Specifically, the question set centred on three core areas namely *preparation and hosting* (venue, information, logistics), *content and delivery* (materials, mock exercises, delivery, group discussions, length of training) and perceived *usefulness and impact*. A full question set is provided in Appendix 1.



As can be seen in Chart 1, above, respondents⁸ were notably positive in their feedback. Logistics were reported as satisfactory or very satisfactory in all cases⁹ with an average rating of 4.86 (out of 5) achieved overall. In terms of practical information, whilst there were a small number of cases where participants talked of difficulties in organising accommodation, remaining respondents were either satisfied or very satisfied with an average rating of 4.58 recorded. In terms of the training venue, there was only one (unexplained) neutral score, with remaining responses positive and achieving an average rating of 4.72. In no case was a negative response recorded for preparation and hosting.

⁸ a total of 34 responses were received from 12 European countries (respondent countries not declared in all cases) with staff from the three core NAs not invited to provide feedback as a consequence of having been directly involved in management, hosting and delivery
⁹ all cases refers to cases where a quantified or qualified answer was given when completing the online survey

2. What the People Say: Content and Delivery (KA1)



A greater variety of (average) scores can be seen in chart 2, yet it is important to note that no individual theme or heading drops below an average rating of 4 thus confirming high overall satisfaction. For training materials, 78% of respondents were very satisfied, the remainder satisfied, achieving an overall average of 4.78 (out of 5). For the two mock exercises, in which assessment scores and assessment criteria were addressed, scores were generally positive yet there were a few cases where either access to materials, or the length of notice provided caused respondents to lower their scores - none scored below neutral - with respective overall averages of 4.28 and 4.50. In relation to the quality of training content, respondents were either satisfied (15.6%) or very satisfied (84.4%) with an average rating of 4.84 recorded. Group discussion fared well and recorded an average score of 4.50, yet training length fared least well recording an average rating of only 4.28, with a number of references to the need for a longer onsite training session. In no case was a negative response recorded in relation to content and delivery.

3. What the People Say: Usefulness and Impact (KA1)

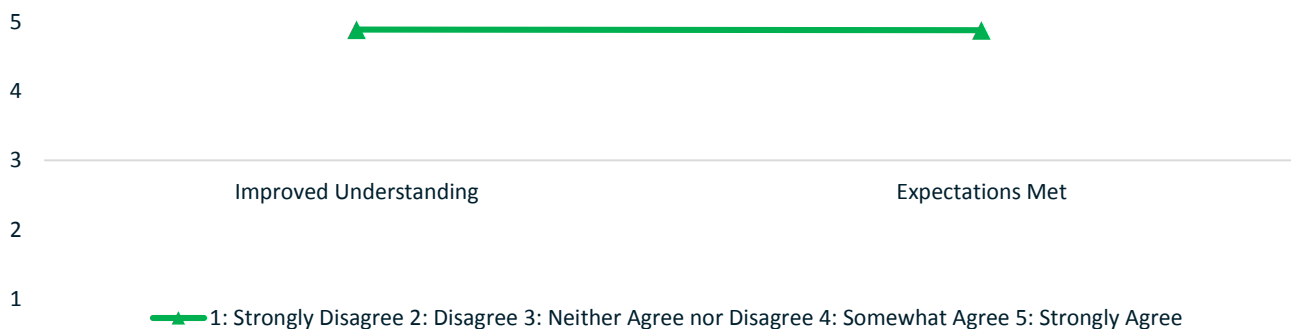


Chart 3 shows the perceived usefulness and impact of the event with questions centred on whether the training improved assessors' understanding of their role and responsibilities and, more generally, on whether event expectations were met. In terms of improved understanding, all respondents agreed (at one level or another) that this was the case, with an overall average rating of 4.89 achieved. In terms of whether expectations were met, respondents were equally positive with overall average rating of 4.88. In no case was a negative response recorded in relation to usefulness and impact yet there was one neutral response given in relation to expectations being met.

Respondents were further asked to provide qualitative input centred on strengths and successes and on recommendations for change and continuous improvement, each separately addressed below.

Perceived **strengths and successes** included the quality and relevance of training materials and the potential for adapting these for use in other training sessions (in the home country) and with other expert groups. The interactive nature of the training sessions was also highly appreciated as it provided an opportunity for the exchange of knowledge and experience and for thematic discussion with peers from other countries, sectors and agencies.

Among those experts recruited to work specifically on the assessment of KA1 applications, a reference was made to improved confidence and enhanced capacity as a result of their participation in the training. Among NA staff, the majority of which also act as assessors for KA1, there was reference to improved understanding of the policies and practices of other agencies (such as interpretations of guidelines and priorities) and to the potential, and inspiration, for influencing change and improvement at their own agencies following their participation in the training.

Examples of comments included:

"having a quiz, instead of a more formal presentation, is an effective way of getting participants' attention"

"I liked the training materials and [having the] chance to exchange experience and common approaches"

"[it] provided an opportunity to understand how others evaluate applications [along with] detailed guidance on how I should approach this work"

"materials and information will be used during training meetings with [other] experts and will also be used by NA staff [when] assessing applications".

In terms of **change and continuous improvement**, there were relatively few suggestions made. The importance of timely and secure access to materials was once again cited - as in 2014 - with a small number of experts not managing to access the "mock exercise" files prior to the day of the event. There were also references to the need for a slightly longer training session, in order for discussions to more fully develop and mature, and for more time to be given to certain exercises such as "budget assessment" and "creating comments".

Examples of comments included:

"I would have liked to receive the material at least a week before the meeting"

"the training should have lasted for at least one more morning or afternoon"

"we had too little time for each topic"

"it would be useful to have more input on how to evaluate the budget... to help decide what to cut".

Expert Training Session for KA2

Following the expert training session in Stockholm (April 2015), which was held over two half days¹⁰, a series of questions were sent to participants with a view to securing input on issues tied to event hosting and event delivery, for which participants were asked to give ratings using a 5-point scale. Specifically, the question set centred on three core areas namely *preparation and hosting* (venue, information, logistics), *content and delivery* (materials, mock exercises, budget assessment, group discussions, trainer, length of training) and perceived *usefulness and impact*. A full question set is provided in Appendix 2.

10 at the end of these two half days, participants were split into two groups focusing on "joint assessment" and "cascading the model"

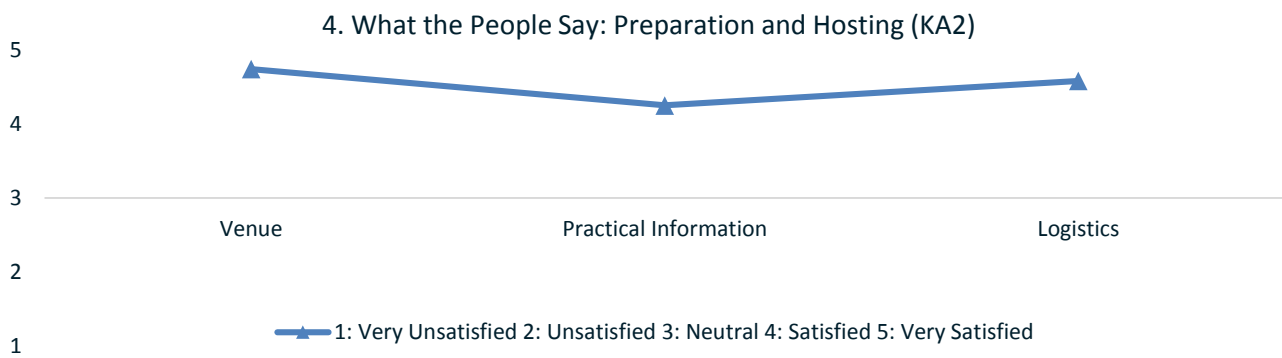
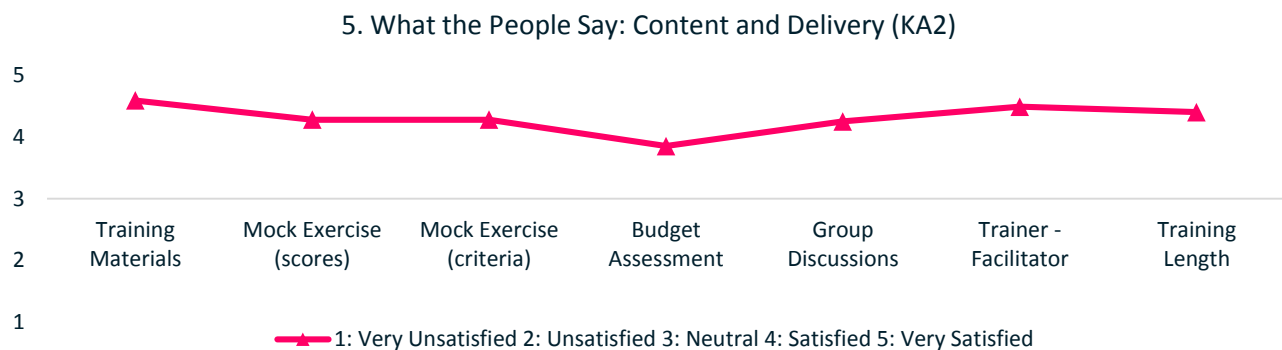


Chart 4, above, confirms that, as a whole, respondents¹¹ were more than satisfied in terms of KA2 event management and hosting. Logistics were predominantly rated as satisfactory or very satisfactory, with an overall average of 4.58 (out of 5) achieved. Practical information scored only slightly lower with an average rating of 4.25 and perspectives on the training venue were positively recorded with an overall average of 4.74. Unlike KA1, however, there were a few cases of “neutral” or “unsatisfied” under the heading of preparation and hosting, with comments centred on the advance mailing of materials (limited notice, messages caught in Spam filters, etc.) and, in one case, to the choice of venue (lack of a Scandinavian feel).

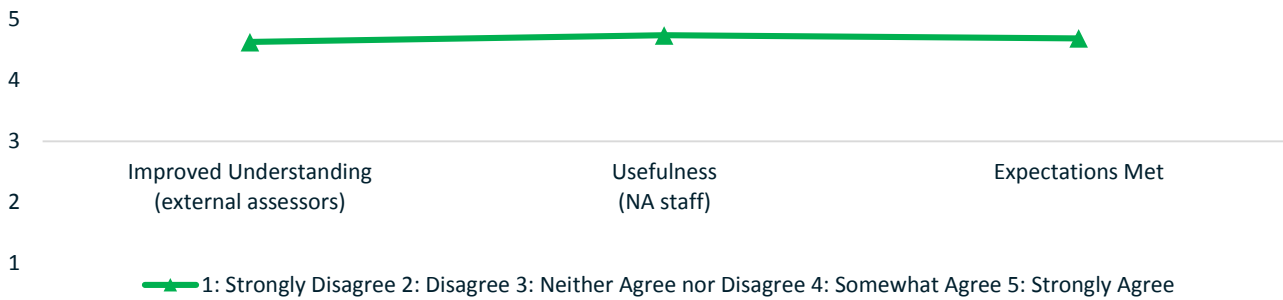
In Chart 5, below, there is one case where the average rating dips slightly below being “satisfied” yet in all other cases, feedback and scores were predominantly positive. For training materials, 94% of respondents were either satisfied or very satisfied, with only a few neutral perspectives recorded and with an overall average rating of 4.59 (out of 5). For the two mock exercises, in which assessment scores and assessment criteria were addressed, scores were generally positive yet there were a few cases where lower ratings were given either as a consequence of having failed to access materials (in advance) or as a reflection of the perceived need for an increased use of new technologies. In one case, it was felt that the trainer-facilitator should have been more directive during the mock exercises, providing preferred scores and perspectives for each of the four assessment criteria.



Least positive among the ratings provided, yet not in any sense negative - having achieved an average rating of 3.85 - is that of budget assessment. In many cases, there were references to the need for more time, greater instruction and/or the use of a real-case budget assessment scenario to enable the required insight to be gained among internal and external experts. Beyond this, ratings revert to positive with an average rating of 4.25 for group discussions (with many comments provided in favour of this practical approach), an average rating of 4.49 for the trainer-facilitator (with few cases of dissatisfaction other than the case where a more directive or instructive input was the preferred model) and with an overall average of 4.40 for the length of the training session (again positive albeit with some calls for slightly longer sessions). There were some negative responses recorded in relation to content and delivery (as previously mentioned) yet, as a whole, these were definitely in the minority with predominantly positive feedback provided.

¹¹ a total of 71 responses were received from 12 European countries (respondent countries not declared in all cases) with staff from the three core NAs not invited to provide feedback as a consequence of having been directly involved in management, hosting and delivery

6. What the People Say: Usefulness and Impact (KA2)



In Chart 6, above, the focus is once again on perceived usefulness and impact with questions centred on whether the training improved understanding among external assessors (of their role and responsibilities), on the perceived usefulness of the event for NA staff working with internal and external assessors and, more generally, on whether event expectations were met. In terms of improved understanding, a large majority of respondents (78.69%) agreed with the statement on increased understanding, with many citing improved confidence and competence and with an overall average rating of 4.63 achieved. Similar levels of satisfaction were recorded among NA staff, with the majority (84.79%) agreeing or strongly agreeing that the training helped in preparing for their own role in assessor training, locally, and with an overall rating of 4.74 achieved. Finally, in terms of the KA2 training event meeting participant expectations, it is positive to see that 95.59% of participants agreed with this statement with an average rating of 4.69 achieved. In no case was a negative response recorded in relation to usefulness and impact.

Qualitative input centred on strengths and successes and on recommendations for change and continuous improvement, each separately addressed below.

There was no shortage of praise for the development and delivery of the KA2 “Model for Expert Training”, with only one negative comment received. Among the cited **strengths and successes** there were direct references to group work and group discussion, to shared perspectives among experts and NA staff, and to the active and interactive nature of training delivery. Others talked very positively of the quality and accessibility of training materials and of the use of practical examples to exemplify assessment goals and practices.

Among those experts recruited specifically to work on the assessment of KA2 applications, there was mention of facilitating assessment through improved understanding, of the value of being able to calibrate scores with other experts and of feeling better equipped (and more confident) to undertake remote assessment. Participating NA staff referred to plans for the continued use of materials (at home) and, in some cases, felt that they had extended their own capacity in terms of being able to confidently respond to questions from expert assessors. Others spoke of plans to deliver similar “interactive” sessions to local experts at some point in the future, and of plans to invite participating (external) experts to relate their experiences to others.

Examples of comments included:

"materials are very good and will probably be translated to our [own] language ... we are also thinking of using the same seminar structure"

"we will use some of the training ideas and methodologies (for example, quiz, fruit basket, red and green cards) as well as some of the materials"

"it helped to [achieve] a more balanced assessment, focusing on what is important and what is not so important"

"it definitely gave me a good insight into what to look for, the tools are especially good and will be used often"

"external experts who took part in the event will pass their experiences to their expert peers".

In terms of **change and continuous improvement**, there were numerous references to the need for more time for budget assessment with some suggesting a real world example be used to solicit expert opinion. In fact, budget assessment scored the lowest among those providing feedback, achieving a mere “satisfactory” rating. Others spoke of the need for the consolidation exercise to be based on a real (or mock) application yet they were equally conscious of time constraints during the common training session. Unsurprisingly, considering the range and number of participants, there were calls for both “less time in plenary” (allowing longer time working in groups) and “more time in plenary” (allowing for a more animated discussion on the work of the different groups). A small number spoke of the potential for making better use of new technologies, including video technologies for presenting some of the more theoretical aspects, and fewer still expressed a desire to work in smaller groups, possibly field-specific. Beyond this, the usual materials access issues were referenced with some suggesting the use of a more well-known platform (for example, Dropbox) to enable access to materials. Finally, there was one request to add numbers to the different training sheets to make them easier to locate during the expert training session.

Examples of comments included:

"[could be useful to have] one live consolidation between two experts so that [participants] could follow the discussion and seek advice from the trainer and from NA staff"

"[could be useful to have] videos online so that we do not have to spend [so much] time lecturing"

"[would have been helpful to have] more input on the budget and the permissible costs"

"[need to have] a longer session on budget categories, particularly intellectual outputs, multiplier events and teaching, training and learning events"

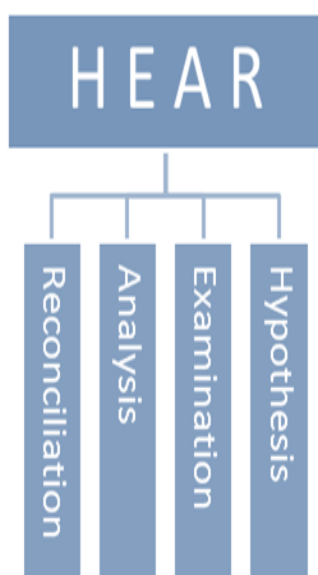
"[I need] more practice in assessing budget calculations"

"[would be useful to have] a bit more time for the conclusions of the working groups during plenary sessions".

In summary, and taking into account feedback received in relation to Expert Training Sessions in Reykjavik (March 2015) and Stockholm (April 2015), it is impossible to ignore the wealth of positivity brought forth by participants. This is clear both in numeric terms - with no single category or topic rated negatively - and in terms of the written feedback that was secured, with proposed changes often adopting an equally positive stance and seeking improvement for the common good.

5. What the Data Says

In this section, a series of *Data Sheets* are provided, each centred on the four-step *HEAR* methodology through which different hypotheses are presented, and considered, alongside data relating to actual assessments undertaken by external experts from Iceland, Norway and Sweden in 2014 and 2015.



Whilst a clear step forward from previous evaluation exercises, in which core data emanated solely from participating (internal and external) experts, it is important to recognise that there are some limits to the current data evaluation exercise, with the following caveats worthy of note:

[a] no baseline or control group data was available, with comparisons reliant on data from the first two years of the Erasmus+ programme, in particular decentralised education and training sub-actions which have a footprint in both 2014 and 2015 - it is expected, however, that as the programmes matures, the validity of the data will increase;

[b] data in this chapter centres solely on three European countries, as opposed to the previous chapter where input was sought from a broader range of participants and countries, with notable differences in population size and with marked differences in the number of applications being submitted and processed - consequently, percentages are most often used, in favour of raw numbers.

As already mentioned, the focus of this section is on those fields (and sub-actions) of education and training which have a footprint in both 2014 and 2015. In this respect, the KA107 and KA200 sub-actions were not included: the former a newly-introduced sub action in 2015, the latter a cross-sectoral action which did not continue beyond 2014. On the contrary, whilst KA219 was newly-introduced in 2015, it represents a composite part of that included previously under KA201 and, as a consequence, data was included. Finally, data relating to mobility actions in the field of Higher Education are not included as a consequence of there being no requirement for qualitative assessment. Data, in all cases, extends to Iceland, Norway and Sweden only.

The table below confirms those actions, fields and sub actions that form a part of the current data review along with the number of applications ultimately assessed, the latter confirming the overall data set as extending to **1151 applications** (602 in 2014; 549 in 2015).

Action	Field	Sub Action	Applications Assessed (2014)	Applications Assessed (2015)
Key Action 1 (KA1)	School Education (SE)	KA101	249	184
	Vocational Education and Training (VET)	KA102	160	133
	Adult Education (AE)	KA104	40	34
Key Action 2 (KA2)	School Education (SE)	KA201	84	20
		KA219	-	84
	Vocational Education and Training (VET)	KA202	24	29
	Higher Education (HE)	KA203	27	37
Adult Education (AE)	KA204	18	28	
TOTALS:			602	549

There are some Data Sheets which focus on convergence or divergence in assessment scoring and, in all such cases, the data set is rather smaller as a consequence of focusing on those applications involving more than one assessor. The table below considers the original data set in terms of whether applications involved one, two or three assessments. With a predominant focus on those applications involving two assessors - the norm in multi-assessor activity, we can see that the reduced data set focuses on **234 applications for KA1** (123 in 2014; 111 in 2015) and on **337 applications for KA2** (148 in 2014; 189 in 2015).

Action	Field	Sub Action	2014				2015			
			Total	1	2	3	Total	1	2	3
KA1	SE	KA101	249	212	35	2	184	155	27	2
	VET	KA102	160	68	84	8	133	51	79	3
	AE	KA104	40	35	4	1	34	29	5	0
Sub-totals:			449	315	123	11	351	235	111	5
KA2	SE	KA201	84	0	82	2	20	0	18	2
		KA219	-	-	-	-	84	0	80	4
	VET	KA202	24	0	21	3	29	0	26	3
	HE	KA203	27	0	27	0	37	0	37	0
	AE	KA204	18	0	18	0	28	0	28	0
Sub-totals:			153	0	148	5	198	0	189	9
TOTALS:			602	315	271	16	549	235	300	14

Finally, it is important to note, when considering analysis and reconciliation of the data, that only in the case of KA2 did the “Model for Expert Training” involve the pooling of experts across the three countries and, as a consequence, in terms of the potential for applications to be assessed by experts from more than one country for KA1, the focus remained on the use of national experts.

Initial data compilation and analysis was undertaken by Paul Guest, expert consultant to the “Model for Expert Training” TCA, with subsequent insights, experiences and perspectives brought forth, by experts from Erasmus+ NAs in Iceland, Norway and Sweden.

In this section, the following twelve topics are addressed in separate Data Sheets:

1. Individual Assessment Scoring (KA1)
2. Individual Assessment Scoring (KA2)
3. Scoring of Assessment Criteria (KA1)
4. Scoring of Assessment Criteria (KA2)
5. Third Assessments by Sub Action (2014 and 2015)
6. Consolidation: Average or Actual (KA1)
7. Consolidation: Average or Actual (KA2)
8. Score Difference by Sub-Action (KA1)
9. Score Difference by Sub-Action (KA2)
10. Scoring and Budgets (KA1)
11. Scoring and Budgets (KA2)
12. Assessment Scoring and Nationality (KA1 and KA2)

Overall findings are presented in Chapter 6: Conclusions.

Data Sheet 1: Individual Assessment Scoring (KA1)

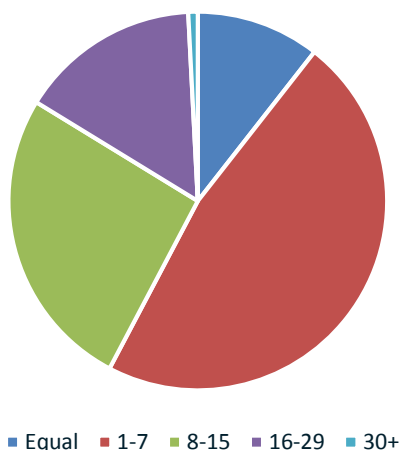
Hypothesis

The “Model for Expert Training” has led to improved capacity among KA1 experts, thus reducing disparities in scoring during individual assessment.

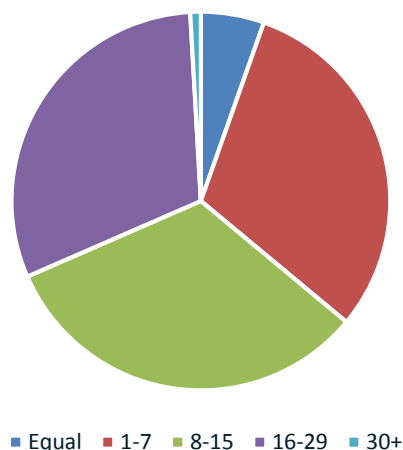
Examination

Score Difference (out of 100) →	Equal Score	1-7 Points	8-15 Points	16-29 Points	>29 Points
2014	10.6%	47.2%	26.0%	15.4%	0.8%
2015	5.4%	30.6%	32.5%	30.6%	0.9%

Differences during Individual Assessment:
KA1-2014



Differences during Individual Assessment:
KA1-2015



Analysis

Data is considered only in cases where two experts were involved during individual assessment - this is non-standard for KA1 and applies to around one third of applications (123 cases in 2014; 111 cases in 2015), specifically those applications requesting higher grant amounts. In categories where the score difference is felt to be negligible or acceptable (≤ 15 percentage points¹²), there is a marked decrease between the years with scores appearing to diverge more in 2015. Applications with a divergence of 30 points or more were very few in both years: 1 in 2014 (0.8% of 123 files) and 1 in 2015 (0.9% of 111 files). Whilst using verified data, internal validity might be affected by one or more additional variables including: a lack of constancy in the expert group (in some cases, new or replacement experts were introduced in 2015), the lack of a baseline or control group and an overall lack of maturity in source data.

Reconciliation

In terms of the hypothesis, whilst it is plausible that the “Model for Expert Training” has improved capacity among KA1 experts, data has yet to fully demonstrate this. It is suggested to repeat this in future years to see if more noticeable trends emerge - possibly using data tied to a constant expert group.

¹² whilst programme rules dictate that a third assessor only needs to be involved once scores differ by 30 points or more, there are often internal rules in place, within different NAs, whereby attention is paid to score divergences of >15 points

Data Sheet 2: Individual Assessment Scoring (KA2)

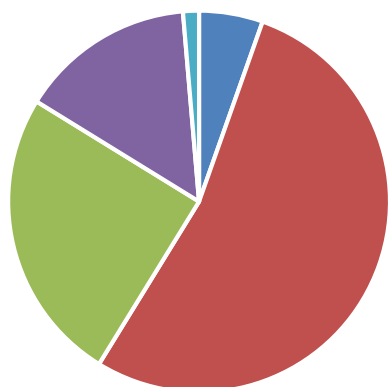
Hypothesis

The “Model for Expert Training” has led to improved capacity among KA2 experts, thus reducing disparities in scoring during individual assessment.

Examination

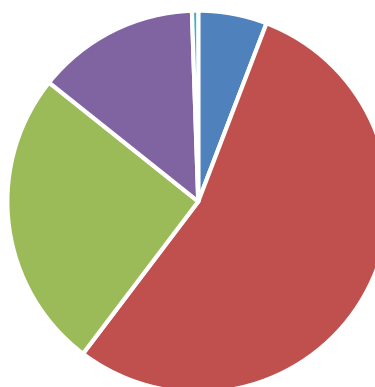
Score Difference (out of 100) →	Equal Score	1-7 Points	8-15 Points	16-29 Points	30+ Points
2014	5.4%	53.4%	25.0%	14.8%	1.4%
2015	5.8%	54.5%	25.4%	13.8%	0.5%

Differences during Individual Assessment:
KA2-2014



■ Equal ■ 1-7 ■ 8-15 ■ 16-29 ■ 30+

Differences during Individual Assessment:
KA2-2015



■ Equal ■ 1-7 ■ 8-15 ■ 16-29 ■ 30+

Analysis

Data is considered only in cases where two experts were involved - this is the standard assessment model for KA2 and extends to 148 (out of 153) cases in 2014 and 189 (out of 198) cases in 2015: remaining cases involved three experts. In categories where the score difference is felt to be negligible or acceptable (≤ 15 percentage points¹³), there is a slight increase in 2015 (1.93%) confirming increased score convergence. There is also a reduction in the amount of applications where there was a divergence of 30 points or more albeit with raw data confirming relatively little change in real numbers: from 2 applications in 2014 (1.4%) to 1 application in 2015 (0.5%). Whilst relying on verified data, internal validity might be affected by one or more additional variables including: a lack of constancy in the expert group (in some cases, new or replacement experts were introduced in 2015), the lack of a baseline or control group and an overall lack of maturity in the source data.

Reconciliation

In terms of the hypothesis, whilst plausible that the “Model for Expert Training” has improved capacity among KA2 experts, it is suggested to look at this in the longer term, with a view to identifying stronger trends - possibly using data tied to a constant expert group.

¹³ whilst programme rules dictate that a third assessor only needs to be involved once scores differ by 30 points or more, there are often internal rules in place, within different NAs, whereby attention is paid to score divergences of >15 points

Data Sheet 3: Scoring of Assessment Criteria (KA1)

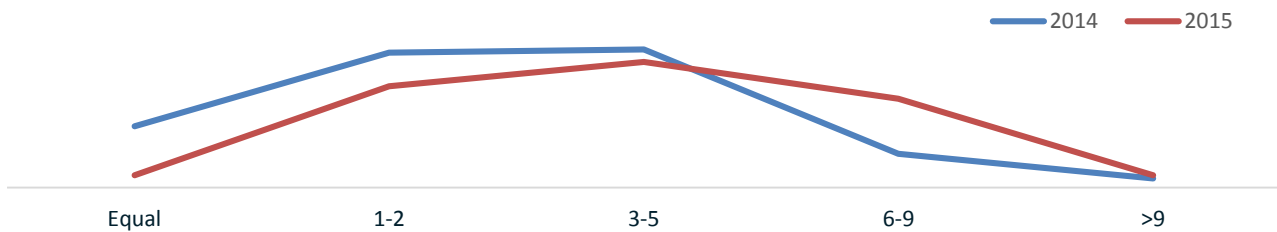
Hypothesis

The “Model for Expert Training” has led to improved capacity among KA1 experts, thus reducing disparities in the scores applied to distinct assessment criteria as a part of individual assessment.

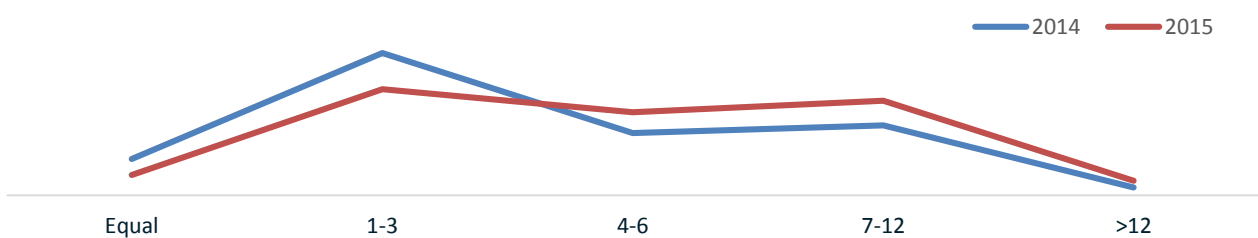
Examination

Score Difference (out of 30) →	Equal Score	1-2 Points	3-5 Points	6-9 Points	>9 Points
Relevance: 2014	16% [20]	36% [44]	37% [45]	9% [11]	2% [3]
Relevance: 2015	4% [4]	30% [33]	37% [41]	26% [29]	3% [4]
Score Difference (out of 40) →	Equal Score	1-2 Points	3-4 Points	5-6 Points	>6 Points
Quality of Project Design: 2014	11% [14]	45% [55]	20% [24]	22% [27]	2% [3]
Quality of Project Design: 2015	6% [7]	33% [37]	26% [29]	30% [33]	5% [5]
Score Difference (out of 30) →	Equal Score	1-2 Points	3-5 Points	6-9 Points	>9 Points
Impact and Dissemination: 2014	14% [17]	39% [48]	33% [41]	6% [7]	8% [10]
Impact and Dissemination: 2015	12% [13]	26% [29]	39% [43]	16% [18]	7% [8]

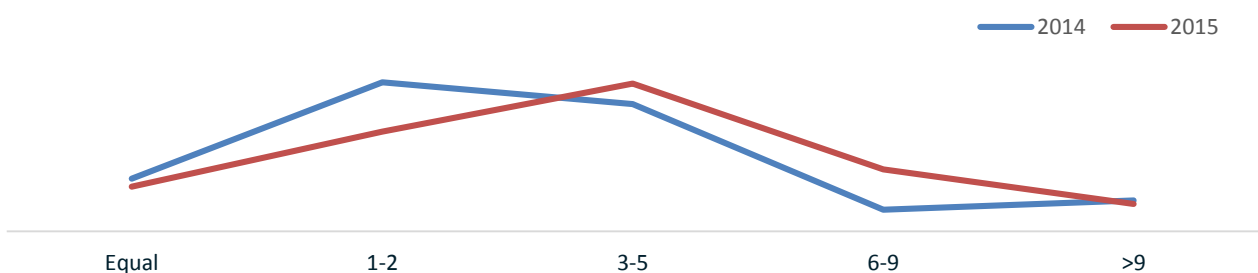
[a] Score Differences out of 30: Relevance (KA1)



[b] Score Differences out of 40: Quality of Project Design (KA1)



[c] Score Differences out of 30: Impact and Dissemination (KA1)



Analysis

Data is considered only in cases where two experts were involved during the individual assessment - this is non-standard for KA1 and applies to around one third of applications (123 cases in 2014; 111 cases in 2015), specifically those requesting higher grant amounts. Divergence for each criterion is considered in terms of the overall score that could be awarded (these differ across the three assessment criteria): for example, a score difference of more than 9 points (out of 30) under relevance would represent a divergence of more than 30% which is the point at which third assessment would be sought in overall scoring terms. For the sake of clarity, data is provided both as percentages and actual numbers (of applications assessed by two experts). For all three assessment criteria, data confirms lesser convergence (greater score difference) in 2015 than in 2014. Under “relevance”, we can see a drop from 89% in 2014, to 70% in 2015. In real terms, this represents 31 fewer cases in which expert scores were close. Under “quality of project design” - arguably the most complex of KA1 assessment criteria with up to 8 different elements expected to be judged - differences are rather more marginal yet also see a drop from 75% convergence in 2014 to 66% convergence in 2015. In real terms this represents 20 fewer cases of expert convergence. Under Impact and Dissemination, we see a drop from 86% in 2014, to 77% in 2015 which, in real terms, represents 21 fewer cases of convergence. Whilst using verified data, internal validity might be affected by one or more additional variables including: a lack of constancy in the expert group (with, in some cases, new or replacement experts introduced), the lack of a baseline or control group and an overall lack of maturity in the source data.

Reconciliation

In terms of the hypothesis, whilst it could be argued that the “Model for Expert Training” has improved capacity among KA1 experts in both years (2014, 2015) with high overall convergence levels, there is no emerging trend that confirms change and improvement between the two years, rather the opposite. It is suggested to gather data over a longer period with a view to confirming longer-term trends.

Data Sheet 4: Scoring of Assessment Criteria (KA2)

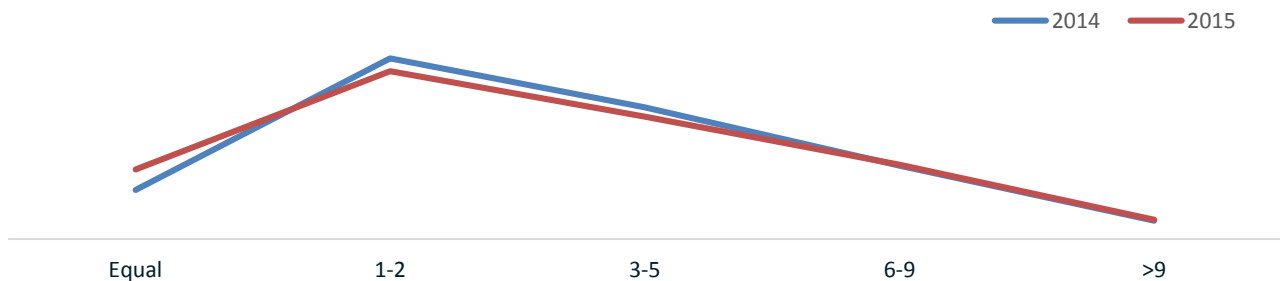
Hypothesis

The “Model for Expert Training” has led to improved capacity among KA2 experts, thus reducing disparities in the scores applied to distinct assessment criteria as a part of individual assessment.

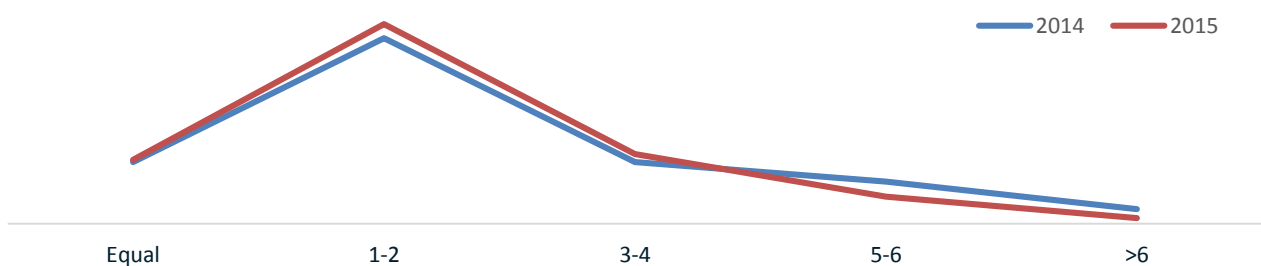
Examination

Score Difference (out of 30) →	Equal Score	1-2 Points	3-5 Points	6-9 Points	>9 Points
Relevance: 2014	11% [16]	40% [59]	29% [43]	16% [24]	4% [6]
Relevance: 2015	15% [29]	37% [70]	27% [51]	17% [31]	4% [8]
Score Difference (out of 20) →	Equal Score	1-2 Points	3-4 Points	5-6 Points	>6 Points
Quality of Project Design: 2014	17% [25]	51% [75]	17% [25]	11% [17]	4% [6]
Quality of Project Design: 2015	18% [33]	54% [103]	19% [36]	7% [14]	2% [3]
Score Difference (out of 20) →	Equal Score	1-2 Points	3-5 Points	6-9 Points	>9 Points
Quality of Project Team: 2014	24% [36]	43% [63]	19% [28]	9% [14]	5% [7]
Quality of Project Team: 2015	22% [42]	51% [96]	18% [34]	7% [13]	2% [4]
Score Difference (out of 30) →	Equal Score	1-2 Points	3-5 Points	6-9 Points	>9 Points
Impact and Dissemination: 2014	21% [31]	42% [62]	22% [32]	12% [18]	3% [5]
Impact and Dissemination: 2015	14% [26]	35% [66]	36% [68]	14% [27]	1% [2]

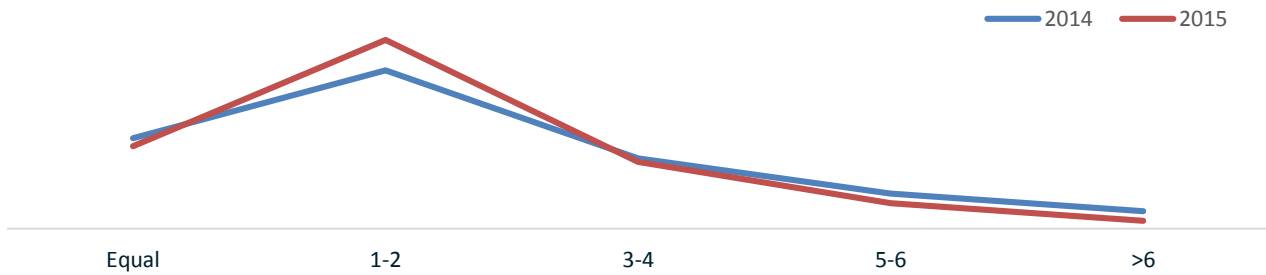
[a] Score Differences out of 30: Relevance (KA2)



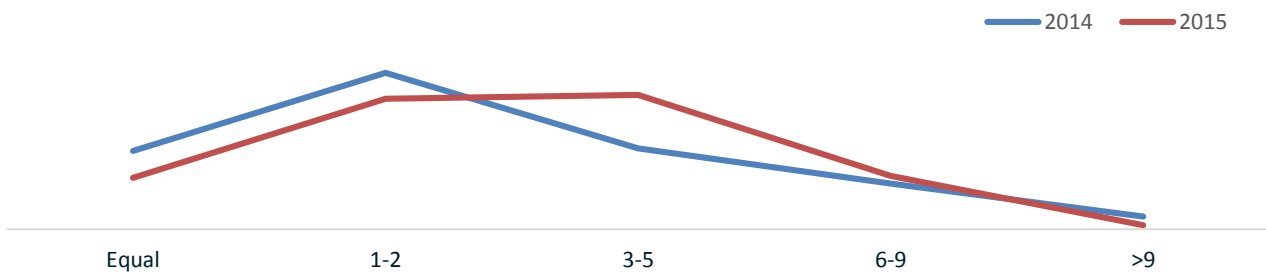
[b] Score Differences out of 20: Quality of Project Design (KA2)



[c] Score Differences out of 20: Quality of Project Team (KA2)



[d] Score Differences out of 30: Impact and Dissemination (KA2)



Analysis

Data is considered only in cases where two experts were involved - this is the standard assessment model for KA2 and extends to 148 (out of 153) cases in 2014 and 189 (out of 198) cases in 2015: remaining cases involved three experts. Divergence for each criterion is considered in terms of the overall score that could be awarded (these differ across the three assessment criteria): for example, a score difference of more than 9 points (out of 30) under relevance would represent a divergence of more than 30% which is the point at which third assessment would be sought in overall scoring terms. For the sake of clarity, data is provided as percentages and as actual numbers (of applications assessed by two experts). Across all four assessment criteria, there is a definite leaning towards convergence among experts, with peaks evident, in almost all cases, at 1-2 points of score difference. An exception is “impact and dissemination” where convergence peaks at 3-4 points of difference (out of 30), not in any way significant. A notable trend is the fact that, in most cases, the higher the difference in expert scores, the lower the number of involved applications. For “relevance”, “quality of project design” and “quality of project team” convergence patterns are extremely similar in 2014 and 2015. Notably, there are visible increases in terms of positive convergence for “quality of project design” (6% increase) and “quality of project team” (5% increase). Whilst using verified data, internal validity might be affected by one or more additional variables including: a lack of constancy in the expert group (with, in some cases, new or replacement experts introduced), the lack of a baseline or control group and an overall lack of maturity in the source data.

Reconciliation

In terms of the hypothesis, it could be argued that the “Model for Expert Training” has improved capacity among KA2 experts with recognisable increases in score convergence among participating experts (2014-2015). Such trends could be more strongly argued where longer term data was gathered and reviewed.

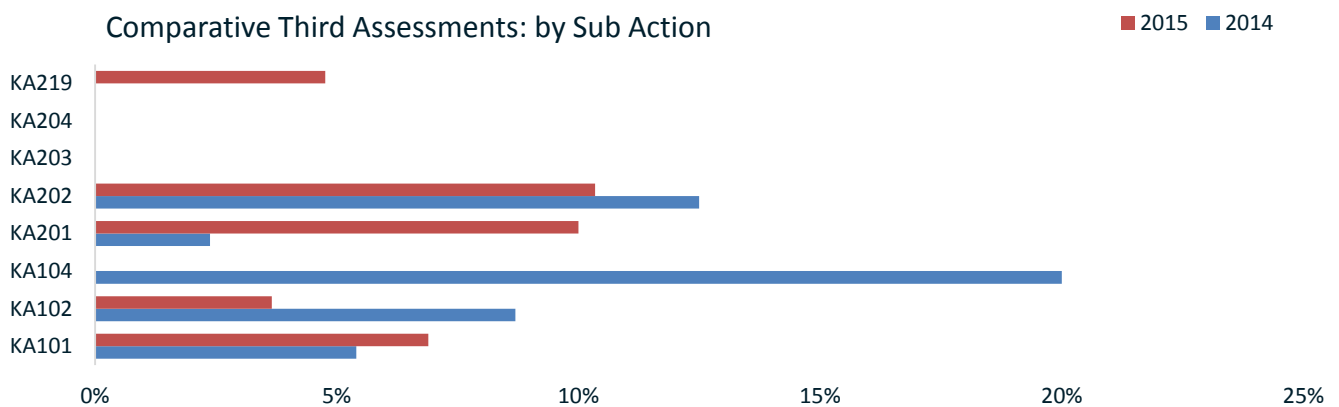
Data Sheet 5: Third Assessments by Sub Action (2014 and 2015)

Hypothesis

The “Model for Expert Training” has led to fewer (individual) third assessments for KA1 and KA2.

Examination

Sub Action →	KA101	KA102	KA104	KA201	KA202	KA203	KA204	KA219
2014	5% [2]	9% [8]	20% [1]	2% [2]	13% [3]	0% [-]	0% [-]	- [-]
2015	7% [2]	4% [3]	0% [-]	10% [2]	10% [3]	0% [-]	0% [-]	5% [4]



Analysis

Data jointly-considers third assessments for KA1 and KA2 (of which there were 30 cases in total) as a percentage of those applications requiring more than one assessment (601 cases), and confirms that this applied to only 5.6% of applications in 2014 and 4.5% of applications in 2015. With such small numbers involved, data is provided both as percentages and as actual numbers (of applications assessed by three experts). In the field of school education, we can see increase in the amount of third assessments for both mobility (KA101: up 2%) and strategic partnerships¹⁴ (KA201/KA219: up 4%). For VET, the opposite is clear with decreases in the number of third assessments for mobility (KA102: down 5%) and strategic partnerships (KA202: down 3%). For higher education, only strategic partnership data is considered¹⁵ with each year positively recording a zero-return in terms of third assessment. For adult education, there are also positive results recorded with a drop from 20% to 0% for mobility projects and with no third assessments recorded in either year for strategic partnership applications. In all cases, it is important to consider that data reflects only small numbers with even a figure of 20% (KA2014: 2014) reflecting third assessment in only 1 out of 5 applications.

Reconciliation

In most cases, data supports the original hypothesis and confirms that fewer third assessments were needed, arguably a consequence of improved understanding of the assessment and consolidation process among experts and, in some cases, a reflection of converging expert opinion. A single exception is school education where data confirms a 100% increase in third assessments (4 in 2014; 8 in 2015).

¹⁴ for strategic partnerships, 2015 data for KA201 and KA219 was jointly-compared to 2014 data for KA201, a consequence of having divided the sub-actions for school education in 2015

¹⁵ no qualitative assessment is undertaken for mobility projects in the field of higher education thus data not considered

Data Sheet 6: Consolidation: Average or Actual (KA1)

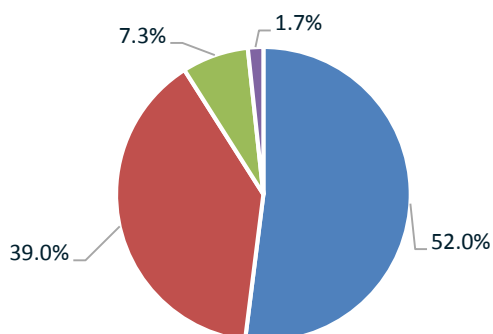
Hypothesis

An increased focus on Consolidation, in 2015, has encouraged KA1 experts to more actively discuss their individual assessments, moving away from the use of simple mathematical averages.

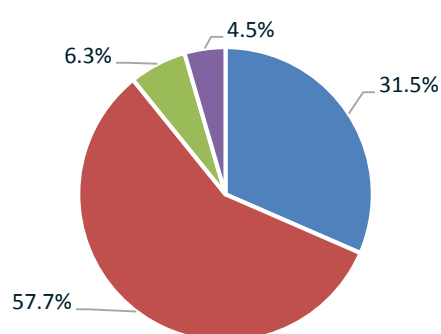
Examination

Difference between "Average" and "Actual" Consolidation Scores →	Mathematical Average Used	Divergence (up to 5 Points)	Divergence (6-10 Points)	Divergence (over 10 Points)
2014	52.0%	39.0%	7.3%	1.7%
2015	31.5%	57.7%	6.3%	4.5%

Consolidation Average or Actual: KA1-2014



Consolidation Average or Actual: KA1-2015



■ Average Used ■ Up to 5 Points ■ 6-10 Points ■ > 10 Points ■ Average Used ■ Up to 5 Points ■ 6-10 Points ■ > 10 Points

Analysis

Data is considered only in cases where two experts were involved during the individual assessment - this is non-standard for KA1 and applies to around one third of applications (123 cases in 2014; 111 cases in 2015), specifically those requesting higher grant amounts. Data confirms reduced reliance on mathematical averages during consolidation (reduction of 18.7%) with the majority of experts, in 2015, agreeing on a consolidated score that sits within 5 points of the mathematical average. Notable divergence from the mathematical average (i.e. 10 points either side) is not significant in either year, which is consistent with the fact that this would only be expected where one expert was particularly enthusiastic (or unenthusiastic) about a particular application and was subsequently able to convince an expert peer of their own rationale for low or high scoring.

Reconciliation

Whilst there is no guarantee that the use of mathematical averages is not the natural result of intra-expert discussion - for example, scores for one assessment criteria might be reduced whereas scores for another criteria might be increased resulting in the same overall score - it is often the case that there will be some divergence from these averages. Data does support the original hypothesis although it is worth noting that in addition to having building additional capacity among participating experts, there was additional financial remuneration given to those experts taking the lead in consolidation in 2015.

Data Sheet 7: Consolidation: Average or Actual (KA2)

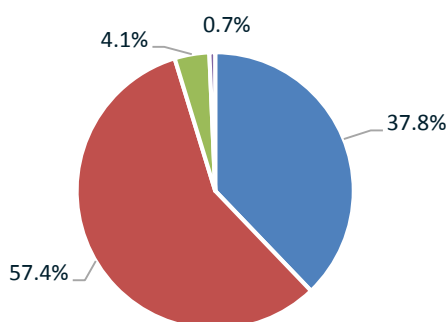
Hypothesis

An increased focus on Consolidation, in 2015, has encouraged KA2 experts to more actively discuss their individual assessments, moving away from the use of simple mathematical averages.

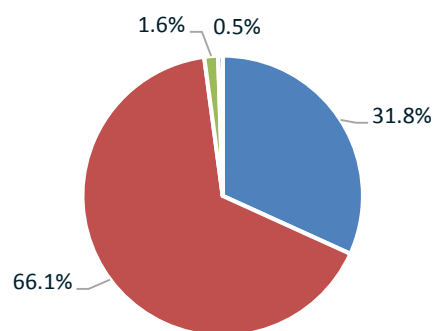
Examination

Difference between "Average" and "Actual" Consolidation Scores →	Mathematical Average Used	Divergence (up to 5 Points)	Divergence (6-10 Points)	Divergence (over 10 Points)
2014	37.8%	57.4%	4.1%	0.7%
2015	31.8%	66.1%	1.6%	0.5%

Consolidation Average or Actual: KA2-2014



Consolidation Average or Actual: KA2-2015



■ Average Used ■ Up to 5 Points ■ 6-10 Points ■ > 10 Points

Analysis

Data is considered only in cases where two experts were involved - this is the standard assessment model for KA2 and extends to 148 (out of 153) cases in 2014 and 189 (out of 198) cases in 2015: remaining cases involved three experts. Data confirms reduced reliance on mathematical averages during consolidation (reduction of 6.0%), albeit not on such a significant scale as in KA1¹⁶, with the majority of experts, in both years, agreeing on a consolidated score that sits within 5 points of the mathematical average. Notable divergence from the mathematical average (i.e. 10 points either side) is not significant in either year (1 case in 2014, 1 case in 2015), which is consistent with the fact that this would only be expected where one expert was particularly enthusiastic (or unenthusiastic) about a particular application and was subsequently able to convince an expert peer of their own rational for low or high scoring.

Reconciliation

Whilst there is no guarantee that the use of mathematical averages is not the natural result of intra-expert discussion - for example, scores for one assessment criteria might be reduced whereas scores for another criteria might be increased resulting in the same overall score - it is often the case that there will be some divergence from these averages. Data supports the original hypothesis although it is worth noting that in addition to having built additional capacity through the strengthened training of experts, there was additional financial remuneration given to experts taking the lead in consolidation in 2015.

¹⁶ findings presented separately in Data Sheet 6

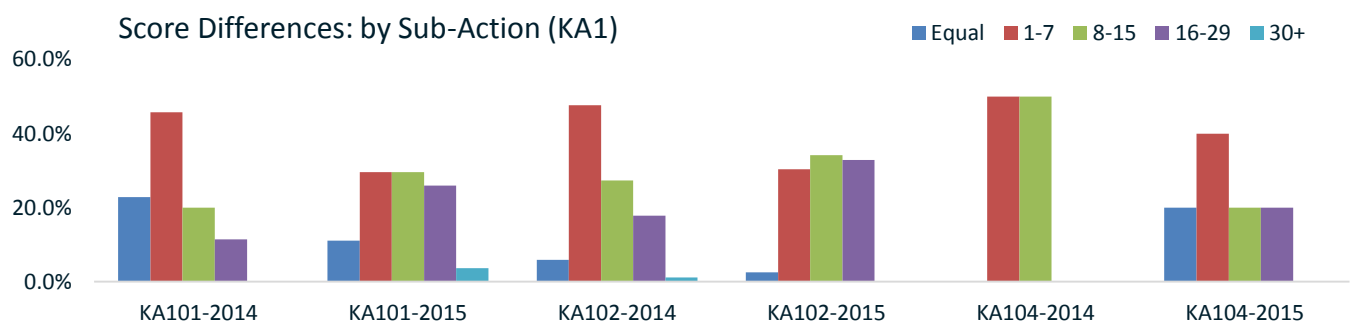
Data Sheet 8: Score Difference by Sub-Action (KA1)

Hypothesis

There are differences in scoring patterns (convergence, divergence) depending on the fields of education and training that are being assessed by KA1 experts.

Examination

Score Difference →	Equal Score	1-7 Points	8-15 Points	16-29 Points	30+ Points
KA101-2014 (SE)	22.9%	45.7%	20.0%	11.4%	0.0%
KA101-2015 (SE)	11.1%	29.6%	29.6%	26.0%	3.7%
KA102-2014 (VET)	6.0%	47.6%	27.4%	17.8%	1.2%
KA102-2015 (VET)	2.5%	30.4%	34.2%	32.9%	0.0%
KA104-2014 (AE)	0.0%	50.0%	50.0%	0.0%	0.0%
KA104-2015 (AE)	20.0%	40.0%	20.0%	20.0%	0.0%



Analysis

Data is considered only in cases where two experts were involved during the individual assessment - this is non-standard for KA1 and applies to around one third of applications (123 cases in 2014; 111 cases in 2015), specifically those requesting higher grant amounts. Data relates only to the fields of school education (SE), vocational education and training (VET) and adult education (AE)¹⁷. Whilst data confirms that there are decreases (rather than increases) in overall expert score convergence¹⁸ there are some notable variances across the different fields and sub-actions. If we consider positive convergence as those cases where scores differed by 15 points or less during individual assessment, we can see that for school education (KA101), there was a clear convergence peak at 1-7 points of difference in 2014 yet with data for 2015 showing increased divergence (i.e. more than 16 points difference out of a maximum score of 100). For VET (KA102), the same increased divergence is evident with positive convergence down from 81.0% in 2014 to 67.1% in 2015. For adult education (KA104), there is a 20% decrease in positive convergence, with previously no significant divergence in this field.

Reconciliation

Data is not conclusive in terms of the initial hypothesis. There is some variance across different fields of education and training yet there is also a general leaning towards positive convergence overall. Notably, there are cases where experts work on more than one field of education and training - arguably worthy of separate review. There is little notable progression between 2014 and 2015 (rather the contrary) with a suggestion that data continues to be gathered with a view to observing longer-term trends.

¹⁷ no qualitative assessment is undertaken for mobility projects in the field of higher education thus data not considered

¹⁸ findings presented separately in to Data Sheet 1

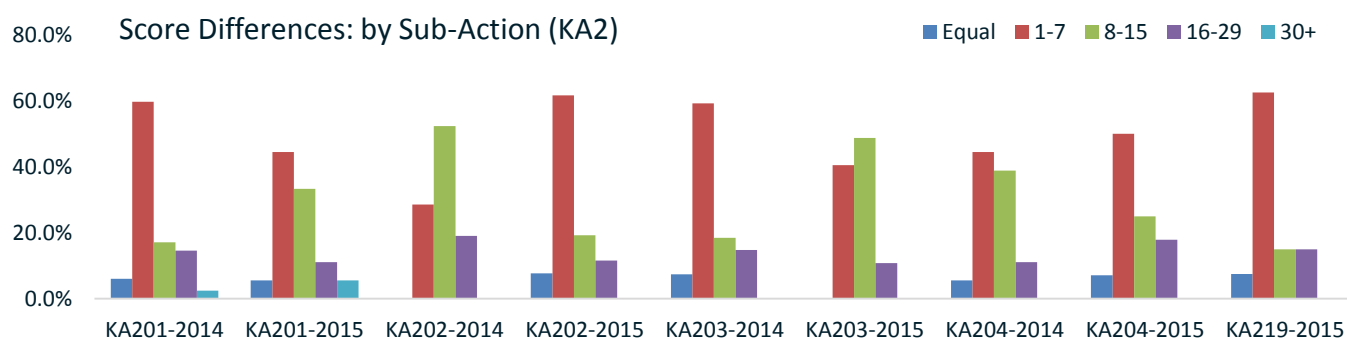
Data Sheet 9: Score Difference by Sub-Action (KA2)

Hypothesis

There are differences in scoring patterns (convergence, divergence) depending on the fields of education and training that are being assessed by KA2 experts.

Examination

Score Difference →	Equal Score	1-7 Points	8-15 Points	16-29 Points	30+ Points
KA201-2014 (SE)	6.1%	59.8%	17.1%	14.6%	2.4%
KA201-2015 (SE)	5.6%	44.4%	33.3%	11.1%	5.6%
KA202-2014 (VET)	0.0%	28.6%	52.4%	19.0%	0.0%
KA202-2015 (VET)	7.7%	61.6%	19.2%	11.5%	0.0%
KA203-2014 (HE)	7.4%	59.3%	18.5%	14.8%	0.0%
KA203-2015 (HE)	0.0%	40.5%	48.7%	10.8%	0.0%
KA204-2014 (AE)	5.6%	44.4%	38.9%	11.1%	0.0%
KA204-2015 (AE)	7.1%	50.0%	25.0%	17.9%	0.0%
KA219-2015 (SE)	7.5%	62.5%	15.0%	15.0%	0.0%



Analysis

Data is considered only in cases where two experts were involved - this is the standard assessment model for KA2 and extends to 148 (out of 153) cases in 2014 and 189 (out of 198) cases in 2015: remaining cases involved three experts. Data relates to all fields of education and training. Whilst data confirms decreases (rather than increases) in overall expert score convergence¹⁹ there are some notable variances across the different fields and sub-actions with some reporting positive improvement, and others not. If we consider positive convergence as those cases where scores differed by 15 points or less during individual assessment, we can see a definite increase in convergence for VET (KA202) and higher education (KA203), with increases of 7.5% and 4.0%, respectively, between 2014 and 2015. For school education, when combining data for KA201 and KA19²⁰, an increase of 1.8% can be observed. On the contrary, for adult education, data for 2015 points towards increased divergence, with 6.8% of applications having a score difference of more than 15 points: in real terms this only represents an increase from 2 applications (of the 18 received in 2014) to 5 applications (of the 28 received in 2015).

¹⁹ findings presented separately in to Data Sheet 2

²⁰ for strategic partnerships, 2015 data for KA201 and KA219 was jointly-compared to 2014 data for KA201, a consequence of having divided the sub-actions for school education in 2015

Reconciliation

Data is not conclusive in terms of the initial hypothesis. Whilst relying on verified data, a lack of constancy in the expert group (in some cases, new or replacement experts were introduced in 2015), the lack of a baseline or control group and an overall lack of maturity in the source data is limiting. There is some variance across different fields of education and training yet there is also a general leaning towards positive convergence overall. Increased convergence is evident in VET and higher education, when considering data for 2014 and 2015, yet data does not confirm whether these are the fields in which experienced - as opposed to new - experts operate. As a whole, it is suggested that data continues to be gathered with a view to observing longer-term trends.

Data Sheet 10: Scoring and Budgets (KA1)

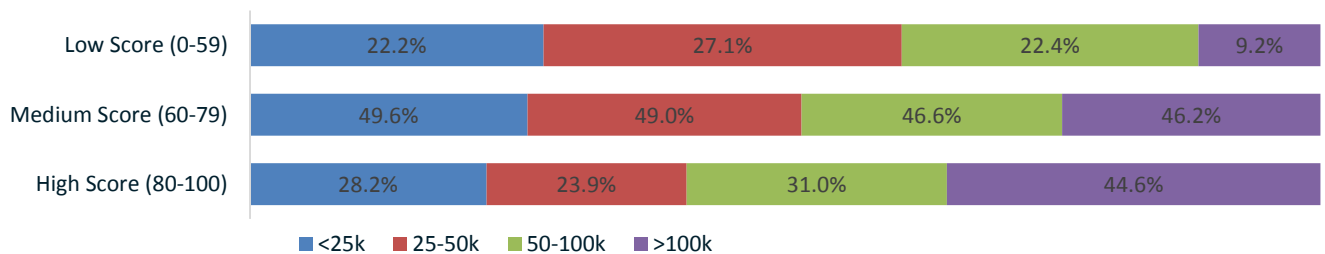
Hypothesis

Experts assessing KA1 applications are likely to award higher scores to larger projects (i.e. those where higher grants are requested).

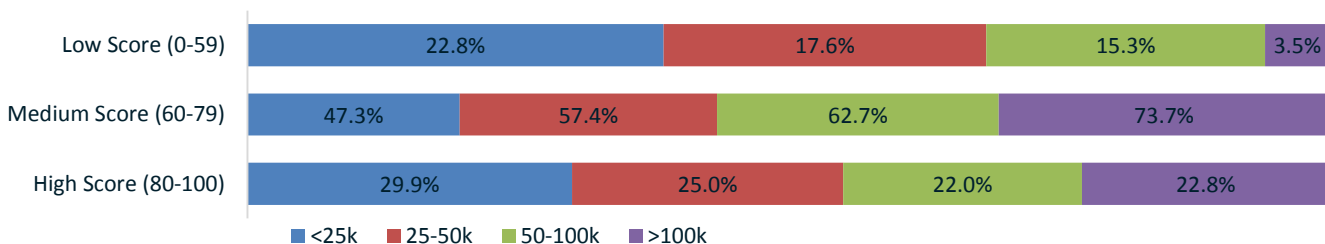
Examination

Grant Requested →	2014				2015			
	<25k	25-50k	50-100k	>100k	<25k	25-50k	50-100k	>100k
Low Score (0-59)	22.2%	27.1%	22.4%	9.2%	22.8%	17.6%	15.3%	3.5%
Medium Score (60-79)	49.6%	49.0%	46.6%	46.2%	47.3%	57.4%	62.7%	73.7%
High Score (80-100)	28.2%	23.9%	31.0%	44.6%	29.9%	25.0%	22.0%	22.8%

Scores and Grant Requested (KA1: 2014)



Scores and Grant Requested (KA1: 2015)



Analysis

Data is considered in relation to all eligible and assessed applications - extending to 800 applications (449 in 2014; 351 in 2015) - and reflects on levels of success across projects in four different funding categories. Data confirms that projects requesting large amounts of funding (>€100k) fared the best in 2014, with 90.8% passing the threshold, higher than any other funding category. In the high scoring band (80-100 points) whilst projects with a larger budget (>100k) achieved the best results in 2014, it was the opposite in 2015 with projects with the smallest budget (<25k) faring best. In terms of passing the overall assessment threshold, and thus being eligible for financing, projects in all categories but one (<25k) were higher in number in 2015.

Reconciliation

Data does not support the hypothesis that a higher budget (or larger project) will achieve greater success in KA1. Figures fluctuate from one year to the next with data for 2014 confirming the attribution of the highest scores to large-scale projects and data for 2015 showing that the highest scores were given to those with the smallest budget. Longer-term data might confirm greater significance for this swing. There is a notable increase in 2015, in terms of the number of projects passing the threshold for financing.

Data Sheet 11: Scoring and Budgets (KA2)

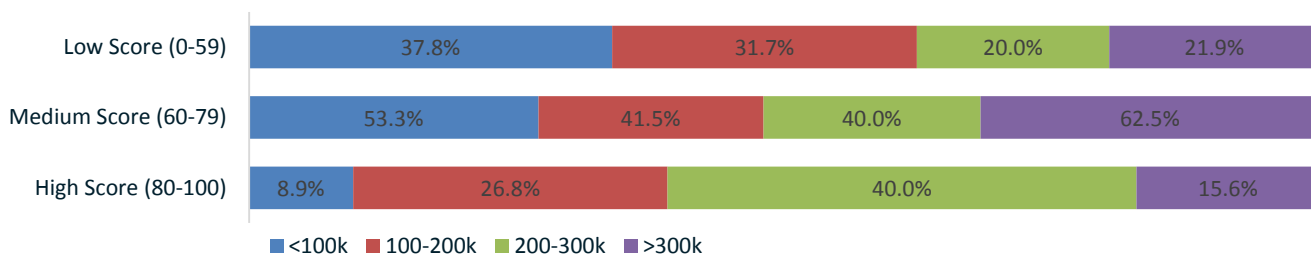
Hypothesis

Experts assessing KA2 applications are likely to award higher scores to larger projects (i.e. those where higher grants are requested).

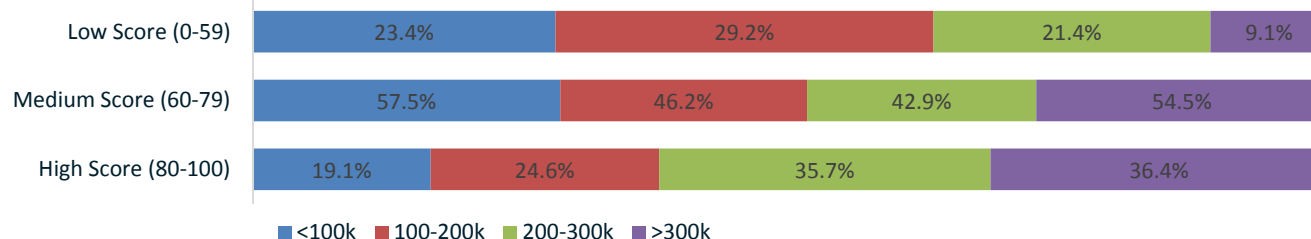
Examination

Grant Requested →	2014				2015			
	<100k	100-200k	200-300k	>300k	<100k	100-200k	200-300k	>300k
Low Score (0-59)	37.8%	31.7%	20.0%	21.9%	23.4%	29.2%	21.4%	9.1%
Medium Score (60-79)	53.3%	41.5%	40.0%	62.5%	57.5%	46.2%	42.9%	54.5%
High Score (80-100)	8.9%	26.8%	40.0%	15.6%	19.1%	24.6%	35.7%	36.4%

Scores and Grant Requested (KA2: 2014)



Scores and Grant Requested (KA2: 2015)



Analysis

Data is considered in relation to all eligible and assessed applications extending to 351 applications (153 in 2014; 198 in 2015) and reflects on levels of success across projects in four different funding categories. In terms of passing the overall assessment threshold, and thus being eligible for financing, most notable are the changes at the two extremes (<100k, >300k) with each showing significant improvement - in terms of the number of projects passing the financing threshold - between 2014 and 2015. Also notable is an increase in the number of projects scoring highest (80-100 points) with, once again, projects at the two extremes showing the most significant level of improvement (increase of 10.2% among those requesting <100k, increase of 20.8% among those requesting >300k).

Reconciliation

Data does not support the hypothesis that the higher the budget (or larger the project) the better chance of achieving the highest scores (80-100 points). Data does suggest, however, an overall improvement in the quality of KA2 applications, and this is most notable at the two funding extremes (<100k, >300k).

Data Sheet 12: Assessment Scoring and Nationality (KA1 and KA2)

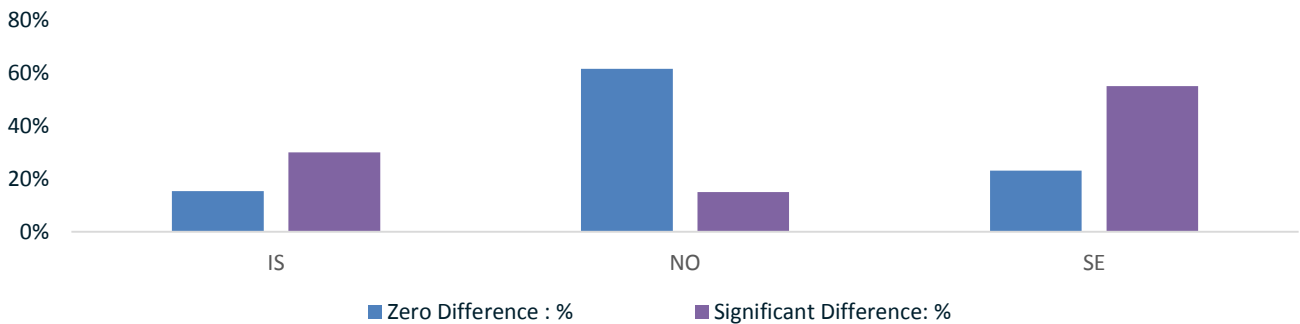
Hypothesis

There are notable differences in scoring patterns (convergence, divergence) depending on the nationality of the expert undertaking the assessment.

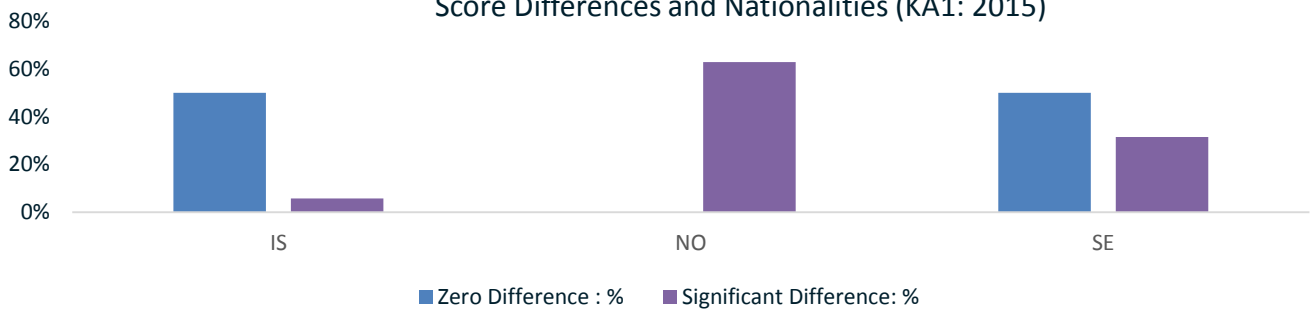
Examination

Action, Country and Year →	KA1						KA2					
	IS		NO		SE		IS		NO		SE	
	2014	2015	2014	2015	2014	2015	2014	2015	2014	2015	2014	2015
Zero Score Difference	15%	50%	62%	0%	23%	50%	25%	14%	25%	32%	50%	54%
Significant Score Difference	30%	6%	15%	63%	55%	31%	4%	11%	54%	35%	52%	54%
Files Attributed	-	-	-	-	-	-	13%	13%	33%	33%	54%	54%

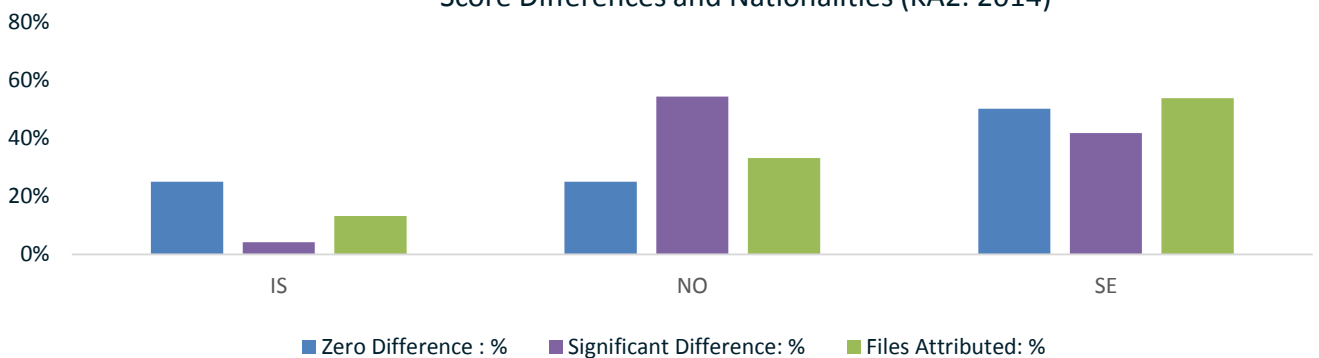
Score Differences and Nationalities (KA1: 2014)

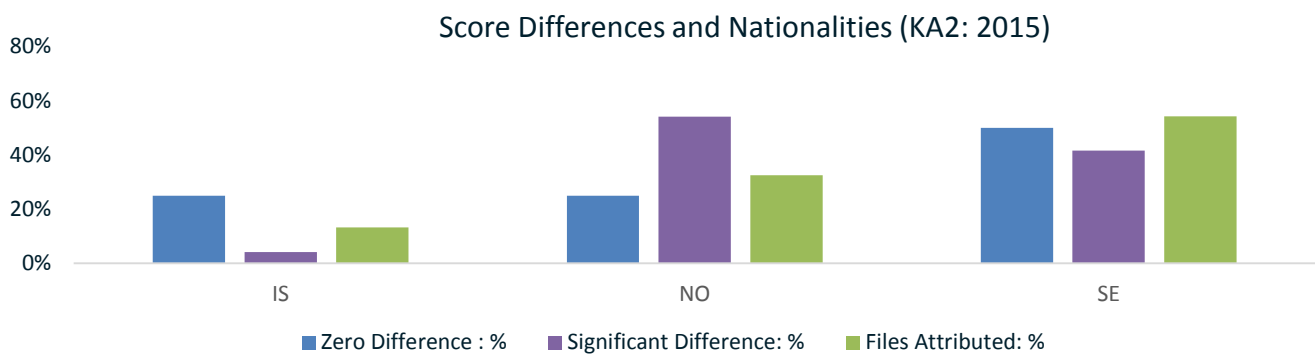


Score Differences and Nationalities (KA1: 2015)



Score Differences and Nationalities (KA2: 2014)





Analysis

Data is considered only in cases where two experts were involved and focuses on two distinct scenarios namely: [a] experts awarding exactly the same score during individual assessment (zero score difference) and [b] experts awarding considerably different scores (>15 points of difference) during individual assessment. Data is considered only for the three core countries (Iceland, Norway and Sweden) and extends to both KA1 and KA2 assessments. It is important to note, however, that KA1 relied solely on national experts during assessment whereas KA2 involved experts from each of the three countries (joint assessment). To allow comparison, the number (%) of assessments that were attributed to experts from each country is also considered for KA2.

Looking initially at KA1, we can see that Norwegian experts had the most cases of “zero difference” in 2014, yet in 2015 there were no such cases for Norway. Looking at raw data for the 26 cases, in 2014, where there was “zero difference”: 4 of these involved Icelandic experts (15%), 6 involved Swedish experts (23%) and the remaining 16 involved Norwegian experts (62%). It is important to remember that these 26 cases reflect only 13 applications, each assessed twice. For 2015, there were 12 cases (6 applications) where there was “zero difference” with an even division between Icelandic experts (50%) and Swedish experts (50%). When looking at “significant difference” in KA1 expert scoring, numbers were definitely higher, increasing from 40 cases (20 applications), in 2014, to 70 cases (35 applications), in 2015. Whilst there is notable progression amongst Icelandic and Swedish experts - the former reducing levels of disparity from 30% (2014) to 6% (2015); the latter reducing from 55% (2014) to 31% (2015) - we have a notable increase in the number of experts having divergent scores for Norway, increasing from 15% (2014) to 63% (2015). Again, figures need to be viewed as a % of those experts featured within the aforementioned scenarios rather than as a % of the whole.

For KA2, patterns for the level of expert involvement in “zero difference” or “significant difference” cases are considered against the depth of involvement of national experts in overall assessment. For example, we can see that Iceland took 13% of the assessment load in 2015, with similar figures presented in terms of “zero difference” (14%) and “significant difference” (11%). The same could be said for Norway where 33% of the assessment load is counterbalanced with 32% “zero difference” and 35% (significant difference). For Sweden, assessment load is 54%, with 2015 data confirming equal figures for “zero difference” and “significant difference”. In all cases, data for 2014 showed greater imbalance suggesting forward progression in terms of expert understanding, in 2015, and resulting in greater convergence.

Reconciliation

Data does not support the initial hypothesis when reflecting over the two years and the two distinct scenarios. There are no obvious trends that favour experts from one or more countries featuring more regularly, or less regularly, in the two predefined scenarios. Contrarily, we can now see greater correlation - when considering figures for KA2 in 2015 - between the percentage of files being assessed by a particular country, and the percentage of assessments that can be categorised within the two distinct scoring scenarios (zero difference, significant difference).

6. Conclusions

From the outset, it is important to acknowledge the dual nature of evaluation activity in 2015, bringing together positive, and occasionally critical, input from those participating in one or more expert training sessions (Chapter 4: What the People Say), alongside analytical reflection centred on data emerging from two years of decentralised assessment activity (Chapter 5: What the Data Says).

Initial evaluation activity built on lessons learned in the past, abandoning onsite evaluation input - often providing little more than an opportunity for self (or peer) appreciation - and using new technologies to seek considered responses, from all participants, in the weeks that followed event attendance. As a whole, overwhelming levels of satisfaction were confirmed among those participating in the expert training sessions for KA1 and KA2 - including in relation to the updated and extended expert training materials - with a number of important suggestions also brought forth in favour of continuous improvement (for example, longer training sessions; easier materials access; use of real case examples; required focus on budget assessment and consolidation). In addition to securing positive feedback from participating experts - confirming improved confidence, competence and understanding - NA participants were equally positive and confirmed the value-added for their own work with experts, highlighting the additional benefits of working together with NA peers.

Beyond this, evaluation activity reflected a first attempt at data analysis and interpretation, centred on 12 distinct hypotheses and covering the first two years of decentralised assessment activity for Erasmus sub-actions in KA1 and KA2. Whilst cognisant of the ambitious nature of such analysis with no obvious baseline data, a relatively immature source data (relying on two years of data) and no control group, a number of interesting messages are already brought forth - for example, whilst joint training has not produced greater harmony in terms of convergence in expert scoring, it has led to a more consensus-based approach where experts target a consolidation of opinion rather than relying on the use of statistical averages; and, whilst data does not support the hypothesis that the higher the budget, the greater chance of achieving a high score during assessment, it does suggest an overall improvement in the quality of applications, most notably under KA2. Furthermore, there is now greater understanding, among the participating (core) NAs, of the value and potential of data analysis and interpretation activity with the majority favouring the extension of this activity to future assessment years, in order to confirm longer-term patterns or trends.

Detailed findings are presented and considered in Chapter 4 and Chapter 5 of this evaluation report. Looking beyond validation of the expert training model, however, and taking into account assessed data alongside feedback gathered from the participating NAs (core and other) in relation to joint expert training and assessment activities, in 2015, the following core conclusions emerge:

- that there is clear value in the joint training of experts with participating experts building competence and confidence and, more importantly, common understanding, enhancing the overall quality of (decentralised) assessment across a range of fields and funding actions;
- that there is clear interest in continued and extended use of the developed expert training model, with many of the participating NAs having confirmed plans for use and delivery in 2015 and beyond;
- that access to experts from other European countries is a positive extension, particularly for smaller countries and in sectors and fields where impartial experts are relatively few, with data confirming equally-high levels of quality in those fields and actions relying on cross-border expertise during assessment;
- that there is definite merit in undertaking continued data analysis and interpretation activity, confirming longer-term patterns and trends - confirming the need for this to become an intrinsic part of future TCA activities centred on the model for expert training.

Appendices

Appendix 1: Question Set for KA1 (QuestionPro)

Appendix 2: Question Set for KA2 (QuestionPro)

Appendix 1: Question Set for KA1 (QuestionPro)

Dear Participant

We thank you for your participation at the Expert training session for Erasmus+ KA1 projects held in Iceland 11-12 March 2015. Please give your feedback on the event, organisation and content.

Thank you in advance.

How satisfied are you with the following:

Preparation and Practical Matters:

	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Satisfied	N/A
Practical information sent prior to the seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Training venue	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logistics during the seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Organisation and Management of the Training Sessions and the Training Material:

	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Satisfied
Training material provided	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mock exercise (sent in advance)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mock exercise (onsite)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of the training content	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of the discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Length of the training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please indicate on the scale how strongly you agree or disagree with the following statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	N/A
The event met my expectations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The training helped me better understand my task as an assessor	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I especially liked/disliked the following aspect/s:

I did not consider the following necessary:

The following question is intended for NA staff only:

How will your organisation benefit from this training?

The following question is intended for expert/assessor only:

How will this training support you in your work?

Comments and suggestions for improvement:

How could the training materials be improved?

How could the training session be improved?

Would you like to make further comments or do you have other suggestions for improvement?

Appendix 2: Question Set for KA2 (QuestionPro)

Dear Participant

Thank you very much for your active participation during the training session for Erasmus+ KA2 projects held in Hasseludden, Sweden, on 19-20 April 2015.

Your opinion of the training model is very important to us so we would very much appreciate if you could take 10 minutes to answer this survey.

Thank you in advance.

The Erasmus+ teams in Iceland, Norway and Sweden.

How satisfied were you with the following:

Preparation and Practical Matters:

	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Satisfied	N/A
Practical information sent prior to the seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Training venue	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Logistics during the seminar	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Organisation and Management of the Training Sessions and the Training Material:

Please tell us how satisfied you were with the following:

	Very Unsatisfied	Unsatisfied	Neutral	Satisfied	Very Satisfied
Training material provided (slides, training sheets)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mock exercise (scores)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Mock exercise (assessment criteria)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Budget assessment session	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Group discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Quality of the trainer	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Length of the training	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Please indicate on the scale how strongly you agree or disagree with the following statements:

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree	N/A
The training event met my expectations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The training helped me better understand my role as an assessor (experts only)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
The training helped me prepare for my role in training own assessors (NA staff only)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

I especially liked/disliked the following aspect/s:

I did not consider the following necessary:

The following question is intended for NA staff only:

How will your organisation benefit from this training in the future?

The following question is intended for expert/assessors only:

How will this training support you in your work as an expert?

Comments and suggestions for improvement:

How could the training materials (training sheets, slides) be improved?

How could the training event or the individual training sessions be improved?

Would you like to make further comments or do you have other suggestions for improvement?