Experiences and lessons from quality evaluations 2011–2014

including follow-ups through 2017

Table of Contents

Introduction

This report summarises the outcome of the cycle of programme evaluations carried out by the Swedish Higher Education Authority (UKÄ) in 2011–2014, and the subsequent follow-ups through April 2017. The report also contains a brief review of the method used during the evaluation cycle, including its background and how it came to be applied to individual evaluations. The report is part of UKÄ's ongoing analysis of its operations. The purpose is to recollect, learn, and disseminate knowledge.

Accordingly, it summarises all experiences and lessons from the evaluations, and their significance to the changes implemented in the new quality assurance system. But the experiences discussed in the report should also be of interest to HEIs in their further development of their own quality assurance systems.

UKÄ has presented an account of experiences from the 2011–2014 evaluation system in previous reports. For example, the report *The Effects of Programme Evaluations* presents and discusses the effects of the evaluations on HEIs' efforts to improve their programmes. The report is based on the measures the HEIs implemented as a result of the evaluations and whether these measures can be specifically linked to the method used during the period.

This report also looks at any effects the method may have had on the outcome of the evaluations, particularly whether the approach could have disadvantaged certain types of programmes. A discussion on this topic arose during the course of the cycle. This discussion is summarised here through a number of questions conveyed in different contexts about the evaluations and linked to potential consequences and effects in the outcome.

Our assessment is that these questions may be considered as general challenges during systematic reviews and evaluations of programme quality. Their examination in the report may thus be of interest to both UKÄ in future reviews and to HEIs when designing their internal quality systems. As the new national system for quality assurance has more sharply focused on HEIs' own quality initiatives, we believe it is important to share our experiences, which may have significance for HEIs' further development of their internal quality systems.

Background, context and questions

Evaluations of the quality of higher education have been carried out in Sweden since the establishment of the National Agency for Higher Education (Högskoleverket, HSV) in 1995.[1] Systematically, this was done during two evaluation cycles. The first cycle comprised six years and was carried out by

---

1. Från granskning och bedömning av kvalitetsarbete till utvärdering av utbildningsresultat – ger utvärderingen en bild av kvaliteten på utbildningen vid universitet och högskolor National Agency for Higher Education 2012:21 R.

HSV from 2001–2006. A total of 1,671 programmes were reviewed at that time, including third-cycle programmes.[2]

In the latest cycle, during which UKÄ assumed responsibility for the task, 2,088[3] programmes were assessed, excluding third-cycle programmes, in just under four years (2011–2014).

In the first cycle, the programmes were evaluated on a two-point scale, and 171 were judged to have inadequate quality. The latest cycle employed a three-point scale and 548 programmes were judged to have inadequate quality, while 298 programmes were judged to have very high quality.
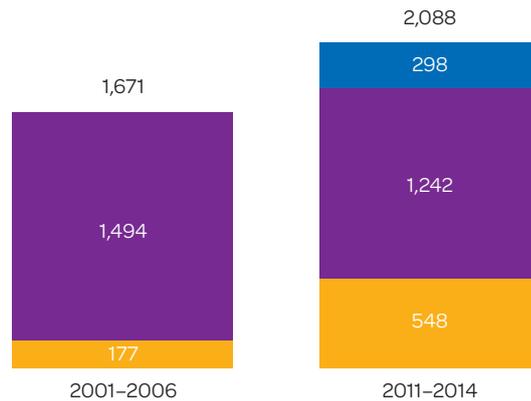
The differences in outcome are clear. This could be explained by the cycles employing different methods, with a focus on examining conditions in the first cycle and examining results in the latest cycle. The methodological differences, which are so extreme that it is difficult to compare the outcomes, are discussed in detail in the report *The Effects of Programme Evaluations*.

The same report concludes that the methods of both systems have been criticised, and several aspects of the critiques are compared and discussed in it. The objections to the latest system are also addressed by former university chancellor Lars Haikola in his review of the method: Assessing Outcomes – Experiences from the Swedish Model.

Indeed, several existing reports discuss its method and effects. What this report brings to the table is a discussion of the actual outcome in relationship to the method and implementation of the evaluations. This discussion is carried out on the basis of a number of questions. Some of these questions refer directly to the 2011–2014 evaluation system, but we also highlight several experiences that are related to evaluations more generally.

---

2. Hur har det gått? En slutrapport om Högskoleverkets kvalitetsgranskningar åren 2001–2006, 2007:31 R.

3. Total of 2,092 programmes if the four programmes at Gotland University College are also included the first round of 2011 evaluations. Because Gotland University College merged with Uppsala University in 2013, these four have been excluded from the summary.

Addressed questions

Snapshot of quality

Did the outcome of the evaluations provide a fair snapshot of the quality of Swedish higher education? This question is particularly relevant with regard to the programmes, i.e., approximately one fourth judged of inadequate quality (a proportion that may seem high, especially compared to the 10 per cent of programmes whose quality was under review during the 2001–2006 cycle).

It is important to note that not all programmes were reviewed in the 2011–2014 cycle. Third-cycle programmes were not included, and several programmes were excluded because the method presupposed a certain number of degree projects. The latter entailed the exclusion of smaller or newly launched or established programmes and qualifications, including teaching programmes and programmes leading to a master of science in business and economics. The number of reviewed first- and second-cycle programmes is noted in an analysis that UKÄ conducted after the cycle's conclusion. The analysis is available through UKÄ's website.[4]

Professional qualifications

How did the heavy emphasis on degree projects impact the evaluation of professional qualifications? Was the method for the 2011–2014 cycle overly focused on theoretical knowledge, and were important components of professional qualifications not taken into account as a result? And what did this mean for the outcome for professional qualifications?

Significance of the number of programmes

Did project size affect how individual programmes were assessed? The varying scope of the evaluations with regard to the number of programmes may influence the outcome. For example, larger evaluations could entail less time permitted per programme and that the assessors conducting the review cannot discuss the individual programmes to the same extent. This question is not primarily about the method used, but is rather a challenge associated with the scope of the evaluation cycle. It resulted in some evaluation projects being very comprehensive. The project size varied from 2 to 85 programmes.[5]

Significance of self-criticism

Are there differences between main fields of study and domains in self-image that systematically influence the outcome of assessments. The point of departure is the assertion that some disciplinary domains/main areas traditionally practice a greater degree of self-criticism than others. This kind

---

4. Kvalitetsutvärderingarna – utfall per lärosäte: http://uka.se/om-oss/publikationer--beslut/statistisk analyser/statistiska-analyser/2015-01-28-uppdaterat-diagram-kvalitetsutvarderingarna---utfall-perlarosate.html.

5. The biggest one consisted of engineering programme assessments, but the size of the projects does not consistently correspond with the number of programmes in the area or qualification, because the engineering programmes were split into clusters/projects containing both professional and general qualifications.

of self-critical approach could result in assessment panels being quick to give the assessed programmes negative judgments when evaluating these areas, even for deficiencies that might be considered non-crucial in other contexts. As a result, when drawing comparisons between assessment outcomes, these areas/domains may appear to be inferior to others when in fact, the outcomes are due to different approaches and academic cultures, rather than to quality. This question may also be relevant in assessments that use other starting points and methods, but UKÄ considers it reasonable to assume that the focus on degree projects may accentuate the effects of these potential differences.

Context

How does one assess unique main fields of study (or specialisations) fairly in systematic evaluations? The context in which a programme is evaluated may be assumed to have significance for how it is assessed. In a results-oriented system where special entry values and conditions are given less weight, it may be more significant. If a programme deviates from the norm, does it risk being criticised for just that or is it given milder treatment because of its differences?

Interdisciplinary programmes

How are interdisciplinary programmes handled in a system with a focus on results? Could the method have disadvantaged interdisciplinary programmes (general qualifications)? This question is also relevant when using other methodologies, but a results-focused method may accentuate the problem. The question involves the context in part, as well as – and perhaps even more so – the fact that the method could be a disadvantage for these programmes due to the focus on immersion in a limited area and insufficient consideration for their more interdisciplinary nature. A similar objection also came up during the 2001–2006 evaluation cycle. At that time, these programmes were potentially at a disadvantage because they were under evaluation with other, more traditionally designed and conventional programmes, subjects with established paradigms.

The concern resulted in a number of programmes being gathered into a special interdisciplinary block towards the end of the cycle. The difficulty of determining in reality which programmes to include in such a block is addressed in the report *Att utvärdera tvärvetenskap*.[6]

The considerations and discussions in this report revolve around these questions. As can be seen above, the questions are not specific to the evaluation system used in 2011–2014, even though they will primarily be discussed based on the outcome of that system. It should also be noted that they were presented in part in less formal contexts. Though some did appear in more formal contexts. For example, several were mentioned in the Government's assignment description for UKÄ in preparation for designing the next system for quality assurance of higher education:

---

6. Att utvärdera tvärvetenskap – reflektioner utifrån Högskoleverkets utvärderingar [Evaluating interdisciplinarity – reflections on the basis of evaluations made by the National Agency for Higher Education] 2001–2005, 2007:34 R.

'Many HEIs have expressed the opinion that the previous system for programme evaluations has mainly been adapted for programmes that lead to the award of a general qualification and is less effective for evaluations of programmes leading to a professional qualification or one in the fine, applied and performing arts. The HEIs have also called attention to the fact that adequate evaluation of many interdisciplinary or multidisciplinary main fields of study has not been possible. Thus there is a strong desire for a national quality assurance system in which different types of programmes can be taken into individual consideration' (Quality Assurance of Higher Education U2015/1626/ UH p. 21).[7]

Finally, it is important to note that the outcome summaries do not aim to compare the quality of programmes in different areas or between HEIs. The evaluations are a snapshot of a moment in time, a measurement that applies to a specific point in time. The data are also not suitable for such comparisons. For example, UKÄ has previously demonstrated the challenges of drawing comparisons between HEIs based on the outcome of evaluations in the analysis *Kvalitetsutvärderingarna – utfall per lärosäte*.[8]

Structure and how to read the report

The structure of the report is based on the questions above. First, a short summary of the method used for the cycle is presented. For a more detailed account of this method, refer to annex 1. Other annexes, which are available on UKÄ's website, www.uka.se: Evaluation Rounds 1–6, Evaluated main fields of study and professional qualifications, Shut-down programmes and withdrawn degree-awarding powers.

Supporting documentation

The report is based on policy documents for the 2011–2014 quality evaluation system, on a selection of evaluation statements and on data from UKÄ's case management system for the management and archiving of UKÄ's assessment activities.

---

7. The opinions in the government assignment derived from, among other places, the main report which summarises the findings of a follow-up carried out by the Danish Centre for Studies in Research and Research Policy on behalf of the Parliamentary Committee on Education, for the purpose of highlighting the implementation and effects of the autonomy reform and quality reform.

8. Diagrams and analysis are available through UKÄ's website, www.uka.se.

Overview: the evaluations and follow-ups

Evaluations

During the last four-year (2011–2014) quality evaluation cycle, the Swedish Higher Education Authority (UKÄ, National Agency for Higher Education through 2012) reviewed a total of 2,088[9] programmes leading to a bachelor's degree, master's degree[10] or professional qualification.

The evaluations were conducted in six rounds, in which the programmes were sorted into clusters that each constituted a project, with a total of 93 projects.[11] In the reviews, each programme was given an overall judgment on a three-point scale. The results of each evaluation were presented in an overall report together with the other programmes in the project.

In accordance with the assignment from the Government of that time, the evaluations during the period focused on assessing the programme outcomes.[12] By 'outcomes', the Government was primarily referring to goal attainment, i.e., whether the programmes achieved the objectives for each qualification as regulated in the Higher Education Ordinance,[13] which stipulates which knowledge, proficiencies and abilities students should have acquired upon completion of a programme.

The assessments of goal attainment were carried out by panels of assessors comprising subject experts, students and labour market representatives. Before each evaluation, the HEI concerned and the relevant professional and student organisations nominate suitable individuals for inclusion on the panel.

The assessment panels judged goal attainment based primarily on three data sets per programme: a selection of student degree projects; a self-evaluation from the HEI supplemented with interviews with programme representatives, and lastly, student interviews.[14]

Based on these data, the panel of assessors determined the goal attainment of each programme under review, i.e., whether the data indicated that the students had achieved the applicable qualitative targets. These assessments by goal were summarised in an overall judgment of the quality of the programme, i.e., an overview of goal attainment. Both the assessment of each individual goal and the overall judgment were given on a three-point scale,

---

9. Total of 2,092 programmes if the four programmes at Gotland University College are also included in the first round of evaluations. Because Gotland University College merged with Uppsala University in 2013, these four have been excluded.

10. The programmes offered at all levels were reviewed individually by level.

11. The plan for the cycle included 88 projects, some of which were split into multiple projects, for a total of 93 projects.

12. Government bill Fokus på kunskap – kvalitet i högre utbildning [Focus on knowledge – quality in higher education].

13. The qualification descriptions in annex 2 to the Higher Education Ordinance.

14. The first evaluations also included an alumni survey.

also pursuant to the government assignment. The choice of terminology for the judgement scale – very high quality, high quality and inadequate quality – was guided by specifications from HEIs that the quality of higher education in Sweden ought to be high.[15] A more detailed description of the method can be found in annex 1.

Follow-ups

For the programmes given the grade of inadequate quality after an evaluation, the HEI had one year to rectify the inadequacies, after which a follow-up was implemented. If the follow-up demonstrated that enough measures had been taken to remedy the inadequacies, the programme was given an overall grade of high quality; in other cases, UKÄ withdrew the relevant degree-awarding power of the HEI.[16] Alternatively, before the follow-up, the HEI opted to shut down the programme. A compilation of the programmes (name of main field of study and professional qualification) that have been shut down or cases in which degree-awarding powers were revoked can be found in annex 4, which is available on UKÄ's website.

---

15. Higher Education Act, chapter 1, section 4.
16. Includes the concepts and forms: university, university college, and higher education provider.

Review of the questions

Snapshot of quality

Did the outcome of the evaluation system provide a fair snapshot of the quality of Swedish higher education with regard to the share of programmes given the grade of inadequate quality?

Educational level

As stated above, 2,088 programmes were reviewed during the 2011–2014 cycle. According to the outcome, 548 programmes were given the grade of inadequate quality, while 298 programmes were given the grade of very high quality. In addition, 1,242 programmes were given the grade of high quality.

This means that 26 per cent received the grade of inadequate quality, 59 per cent received the grade of high quality, and 14 per cent received the grade of very high quality.

Distribution

■ Inadequate quality ■ High quality ■ Very high quality
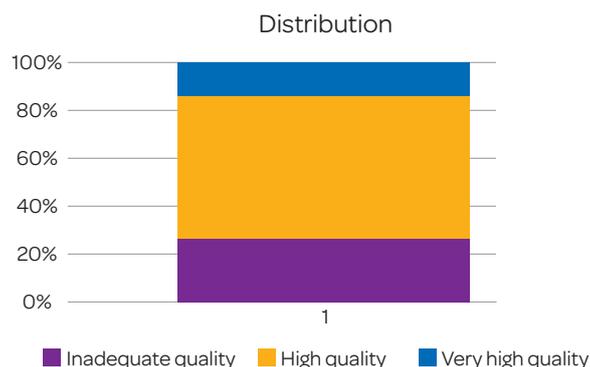
The fact that one fourth of the evaluated programmes received the grade of inadequate quality may seem like a high proportion, but adding the follow-up results paints a different picture. The 548 programmes judged to be of inadequate quality were followed up one year later. Of these, 466 received the grade of high quality in the follow-ups. Of the remaining programmes, 71 were shut down by the HEIs themselves and UKÄ withdrew the degree-awarding powers of the last 10 programmes.[17]

Thus, as a consequence of the evaluations, student admissions have ceased for 81 programmes, corresponding to less than 4 per cent of the 2,088 programmes (which is difficult to illustrate in a diagram). It is worth mentioning that, by design, follow-ups are only carried out for the programmes given the grade of inadequate quality, i.e., those judged not to be high quality according to the legal requirements for this.

From this perspective, it is perhaps less surprising that one fourth of the evaluated programmes were followed up and were expected to present

---

17. The eleventh case involves an independent higher education provider. UKÄ recommended to the Government that the relevant degree-awarding power be withdrawn.

measures taken as a result of problems that were found during the reviews. In the previously implemented cycle (2001–2006), all reviewed programmes had to report on measures taken as a result of the evaluations one year after the report was published.

Another aspect worth mentioning in this context is the significance of the terminology used when the judgement was that the requirements had not been met, i.e. inadequate quality. The logic behind the word choice for the used model was straightforward: inadequate goal attainment resulted in inadequate quality. But the word choice may have contributed to the system being perceived as misleading. The use of 'inadequate quality' has been perceived as simplistic for two primary reasons. First, it did not take into account the scope of the inadequacies, i.e. what they pertained to (which qualitative targets) and how significant they were (one or more qualitative targets, with minor or major inadequacies). Second, the judgement appeared to be final when in reality, the HEI had one year to remedy the inadequacies.

Both the significance of the evaluation follow-up and the terminology of the judgements, including the scale of the latter, were discussed before the launch of the new quality system. The new system aims to make the follow-up part of the learning and development process, to which the reviews are meant to contribute. What this means specifically with regard to whether all programmes or only a selection should be followed up is being tested in the framework of pilot reviews and first rounds to find the most suitable course of action.

The matter of terminology has been easier to handle before the launch of the new system. The change that has been implemented has abandoned 'inadequate quality' in favour of 'programme is under review', which better indicates that there is room for improvement within a given time period. Still, the modified terminology does not resolve the challenge posed by the fact that even moving forward, the judgements will occasionally require nuance with regard to the scope of strengths and weaknesses. But it is worth pointing out that a basic principle remains: when quality is assessed as inadequate, only one judgement remains that demonstrates this, regardless of the scope of the inadequacies or title of the judgement.

Goal level

Yet another perspective on the outcome can be gained if goal attainment is assessed per goal. A total of 11,144 assessments of goal attainment for the 2,088 programmes evaluated were made. The distribution of judgements (1 = inadequate goal attainment, 2 = high goal attainment, 3 = very high goal attainment) for these assessments is illustrated below:

Assessments of goal attainment

Assessments of goal attainment
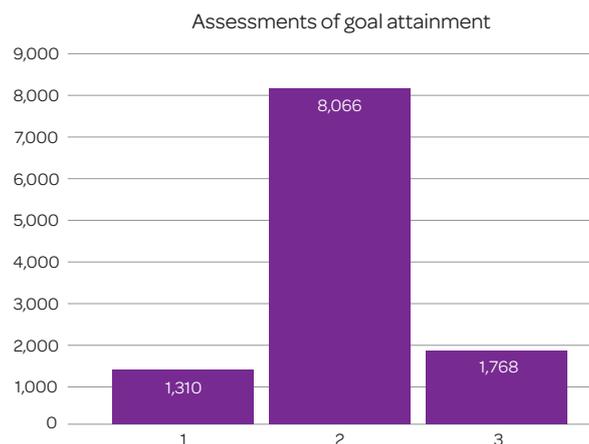
Assessments of inadequate goal attainment comprised about 12 per cent of the total number of assessments, while the share of assessments of very high goal attainment comprised about 16 per cent of the total. Once again, the explanation for the breakdown of inadequacies in relation to the breakdown of very high goal attainment regarding the results for each programme may be credited to the method, and in this case, to the model for the overall judgement for each programme. Inadequate fulfilment of individual qualitative targets meant the programme received an overall judgement of inadequate quality. The method and terminology did not discern between whether the inadequacies were in assuring completion of one or several qualitative targets. On the other hand, an overall judgement of very high quality required an assessment of very high attainment of at least half of the reviewed goals (selection of goals).

The consequences of this are also evident in UKÄ's latest model for evaluations, where the emphasis is now on overall assessments rather than the breakdown or outcome per reviewed component. The goal is to make overall assessments, where the scope of inadequacies in the component should be able to be seen in relationship to the quality of the whole. In addition, there is no judgement for excellence or very high quality; rather, the statements will permit more space for the emphasis of special strengths.

Professional qualifications

How did the heavy emphasis on degree projects impact the evaluation of professional qualifications? Was the method overly focused on theoretical rather than practical knowledge, and if so, did it disadvantage the evaluated professional qualifications?

The table below presents the outcome, divided into professional qualifications and general qualifications including the fine, applied and performing arts. The difference in the number of programmes is great, but the indication is that professional qualifications fared worse.

The degree	Total sum	High quality	Very high quality	Inadequate quality
General (incl. arts)	1,616	972 (60%)	249 (15%)	395 (24%)
Professional alifications	472	270 (57%)	49 (10%)	153 (32%)
Total sum	**2 088**	**1,242 (59%)**	**298 (14%)**	**548 (26%)**

If general qualifications are divided into the various relevant qualification levels, the outcome is as follows:

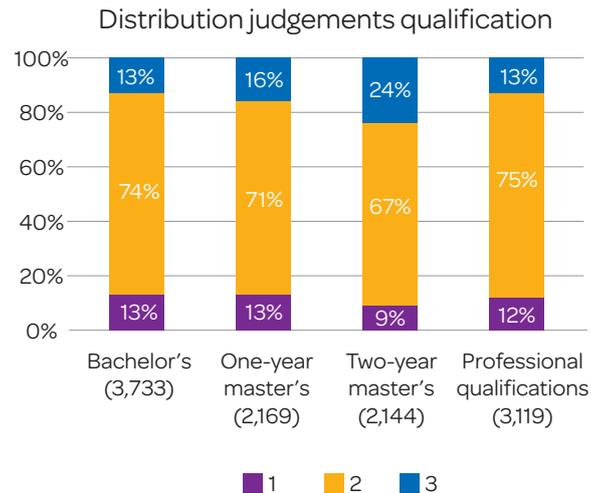| The degree	Total sum	High quality	Very high quality	Inadequate quality
Bachelor's	843	532 (63%)	105 (12%)	206 (24%)
One‑year master's	394	226 (57%)	56 (14%)	112 (28%)
Two‑year master's	379	214 (56%)	88 (23%)	77 (20%)
Professional alifications	472	270 (57%)	49 (10%)	153 (32%)
Total sum	**2,088**	**1,242 (59%)**	**298 (14%)**	**548 (26%)**

Professional qualifications are still worse off, but the division also shows the differences between the levels for general qualifications. The two-year master's degree has the best outcome. Incidentally, this is likely due to the fact that students admitted to two-year master's degree programmes are generally highly motivated, having already completed a bachelor's degree, and are thus well prepared to complete the degree projects that comprise an important part of the two years of studies.[18] When it comes to the outcome for professional qualifications in the 2011–2014 evaluation system, however, the explanation is not as obvious. There are likely several contributing factors.

One reason for the outcome for professional qualifications may be the principle applied to the evaluations to determine the overall quality judgement. This meant that inadequacies in the fulfilment of qualitative targets resulted in inadequate quality, while very high quality required very high fulfilment of at least half of the selected goals.

 Because most professional qualifications have more qualitative targets than general qualifications and more objectives per programme were thus included in the reviewed selection, there is likely an increased risk of discovering inadequacies. At the same time, more objectives make it harder to earn the judgement of very high goal attainment.

This explanation is doubtless supported by the division of the total number of assessments of goal attainment by general qualification level and professional qualification reported below. The figure indicates that beyond two-year master's programmes, the division is fairly similar for the various qualifications. Indeed, there have not been more goal attainment inadequacies among professional qualifications. The fact that the inadequacies had a greater impact on the aggregate level may be because the individual inadequacies in goal attainment for specific goals for professional qualifications had a large impact due to the model for overall judgement.

18. In general, established HEIs have the power to award two-year master's degrees, which attract motivated national and international students and which likely benefits quality at these HEIs.

Distribution judgements qualification

	Bachelor's (3,733)	One-year master's (2,169)	Two-year master's (2,144)	Professional qualifications (3,119)
3	13%	16%	24%	13%
2	74%	71%	67%	75%
1	13%	13%	9%	12%

■ 1 ■ 2 ■ 3

Of the 12 per cent judged to have inadequate goal attainment for professional qualifications, more than half of the evaluations apply to specialist nursing programmes.[19] Together with engineering, they comprise the dominant portion of the number of professional qualifications. Engineering comprises over one third and specialist nursing comprises over one fifth of all programmes.[20] In the specialist nursing evaluation, around 70 per cent of the programmes were rated as having inadequate quality and only five per cent were rated as having very high quality. The impact of the evaluation for the overall outcome for professional qualifications has thus been significant.

Two questions that arise are which qualitative targets the inadequacies applied to and on what basis the judgement of goal attainment was made. The discussion of the system has maintained that the model of reviewing degree projects is less suited to professional qualifications, and judging degree projects therefore disadvantages professional qualifications. The argument has two components, even if each component is not separate from the other.

One component of the argument asserts that degree projects are valued differently in programmes leading to professional qualifications from those leading to general qualifications. This argument implies that degree projects are less important for a professional qualification than for a general qualification. This conclusion lacks a formal basis, as support for it cannot be derived from the Qualification Ordinance, which does not lay out the differences in degree projects between professional and general qualifications.[21] However, perhaps there has traditionally been a difference in the practical design and focus between different qualification categories, where degree projects in professionally oriented qualifications are not allocated the same value as in general qualifications.

---

19. Total of 372 judgements of inadequate goal attainment, of which 197 were for a postgraduate diploma in specialist nursing, i.e., 53 per cent.

20. Total of 472 professional qualifications, of which 105 were for specialist nursing, 94 were for the degree of bachelor of science in engineering, and 78 were for the degree of master of science in engineering.

21. The scope of the degree projects is specified for all professional qualifications in relation to qualification level, except for a teacher qualification for independent adult education and a higher education diploma in vocational education, where no scope is provided.

The other component of the argument relates to many qualitative targets for professional qualifications being considered less suitable for examination through degree projects. It follows that goal attainment for professional qualifications thus cannot be judged via these projects either. The objection is reasonable in light of the fact that professional qualifications have more qualitative targets, many of which may be characterised as programme-specific and experience-related. This objection, however, is based on the perception that degree projects have been the sole basis for the outcome. This has not been the case.

In each evaluation, after choosing the qualitative targets, the experts considered which of the various data sets could comprise the primary basis for assessing whether the programme ensures students' achievement of the respective qualitative targets. This led to some variation between the evaluations based on which basis the assessments were made. Thus, it has differed for different qualifications, even between different professional qualifications.

A review of a selection of reports carried out by UKÄ shows that in evaluations of professional qualifications, the self-evaluation has had major significance for the assessment of students' goal attainment for certain qualifications. The review shows that self-evaluations have had the most significance on the assessment of goal attainment in healthcare qualifications. For qualifications in engineering, self-evaluations have had little significance for the assessment of goal attainment.

The evaluation of the postgraduate diploma in specialist nursing is one example of how degree projects were not the only decisive factor for the outcome. In that evaluation, the assessment panel initially determined that for half of the six relevant qualitative targets, it would be difficult to ascertain goal attainment based on degree projects. Goal attainment for these experience-related objectives would therefore primarily be assessed based on the self-evaluation and interviews. And many of the programmes given the overall grade of inadequate quality in the evaluation were only insufficient in these three experience-related objectives – sometimes in one, but usually in several – where the assessment was indeed made using data other than the degree projects. Of the 105 specialist nursing programmes evaluated, 73 were given the grade of inadequate quality.

Of these 73 programmes, 19 were only inadequate on the basis of self-evaluations and interviews, i.e. 26 per cent (approx. one fourth). Of the total number of programmes included in the evaluation, 18 per cent were only inadequate on the basis of the self-evaluation (about one fifth).

The postgraduate diploma in specialist nursing shows that assessments through data other than degree projects were made and were significant in the 2011–2014 evaluation system. At the same time, it was within the evaluation, and also generally within the system, that degree projects were typically crucial for the outcome. The explanation for this is that students' degree projects were – and probably are – the most suitable basis for drawing conclusions on the results of the programmes based on goal attainment. At any rate, this is the case with the data used thus far in the programme evaluations. And it applies regardless of qualification. Degree projects are not, however, equally suited to all qualitative targets. This conclusion is

important to consider for future evaluations, where programme results are still a component for assessment.

In light of these experiences, UKÄ is now testing ways to balance and weigh together different data and simultaneously find methods for how students' degree projects can also be used in future evaluations of professional qualifications. Methods for this are being tested in the framework of the pilot study for the evaluation of teacher education programmes, which is expected to be complete in autumn 2017.

Significance of the number of programmes

Do the varying sizes of the evaluations regarding the number of programmes have an influence on outcome? Have larger evaluations resulted in less time being permitted per programme and in the assessors conducting the reviews being unable to discuss individual programmes in as much detail?

In this cycle, evaluations were carried out in six rounds. Each round included 10–15 evaluation projects, which the programmes were distributed in; in total there were 93 projects for the 2,088 programmes.[22] This structure was based on an initial mapping of the HEIs' programmes in the relevant qualification levels. Based on this mapping, HSV/UKÄ developed a proposal for the structure of the implementation of the 4-year cycle of evaluations. Before settling on the structure, the HEIs were also given the opportunity to comment on it.

The outcome by evaluation project is presented below, sorted by size. When considering the number, it is important to note that for general qualifications, programmes are included at three qualification levels: bachelor's degree, one-year master's degree and two-year master's degree. It is also worth noting that the title of each evaluation project is generally analogous to the most frequently occurring main field of study in the project, but similar main fields of study are also included.

	Number			Share %		
Evaluation	Number of pro-grammes	Inadequate quality	High quality	Very high quality	Percentage inadequate	Percentage high quality	Percentage very high quality
Data / IT / media technology	85	31	39	15	36%	46%	18%
Mechanical engineering	83	25	50	8	30%	60%	10%
Business administration	65	24	33	8	37%	51%	12%
Engineering physics and electrical engineering	61	12	44	5	20%	72%	8%
Education, didactics and educational sciences	57	18	34	5	32%	60%	9%
Bio / Chemical / Environmental / Energy engineering	50	12	24	14	24%	48%	28%
Urban management	50	12	34	4	24%	68%	8%
Postgraduate diploma in specialist nursing - AMB, AN, INT, OP	49	32	16	1	65%	33%	2%
Earth sciences and human geography	46	5	25	16	11%	54%	35%
Social work and bachelor of science in social work	46	15	30	1	33%	65%	2%
Economics	45	14	29	2	31%	64%	4%
Sociology	42	13	21	8	31%	50%	19%

22. The plan for the cycle indicates 88 clusters. Three of them were divided into additional clusters, of which 93 are presented here. The projects began in six different rounds, divided over the cycle, which formally took place from January 2011 through October 2014.

Biology	42	6	28	8	14%	67%	19%
Political science etc.	41	9	27	5	22%	66%	12%
Film, fashion, art, music, performing art studies	41	8	26	7	20%	63%	17%
Psychology	40	6	27	7	15%	68%	18%
Environmental sciences	39	8	28	3	21%	72%	8%
Religious studies and theology	38	9	17	12	24%	45%	32%
Industrial economics	35	18	13	4	51%	37%	11%
Law	33	7	18	8	21%	55%	24%
English	33	14	13	6	42%	39%	18%
Postgraduate diploma in specialist nursing - PED, DSC	31	20	8	3	65%	26%	10%
Mathematics and mathematical statistics	30	5	12	13	17%	40%	43%
History	30	3	25	2	10%	83%	7%
Chemistry	29	7	21	1	24%	72%	3%
Nursing / health sciences - second - cycle level	29	5	22	2	17%	76%	7%
Media and design	27	5	18	4	19%	67%	15%
Media and Communications	26	7	18	1	27%	69%	4%
Dentistry	26	11	14	1	42%	54%	4%
Informatics	25	2	20	3	8%	80%	12%
Literary studies	25	5	14	6	20%	56%	24%
Swedish and Nordic languages	25	3	15	7	12%	60%	28%
Bachelor's degree nursing / health sciences	25	5	19	1	20%	76%	4%
Postgraduate diploma in specialist nursing - PSYCH, GER, SUR, ONC	25	21	3	1	84%	12%	4%
Special needs ed., special needs training, special ed. needs	25	8	17	0	32%	68%	0%
Physiotherapy	24	0	6	18	0%	25%	75%
Architecture	24	0	23	1	0%	96%	4%
Bachelor of Science in Nursing	24	2	21	1	8%	88%	4%
Peace and conflict studies etc.	23	7	14	2	30%	61%	9%
Public health	23	7	13	3	30%	57%	13%
Romance languages	21	1	16	4	5%	76%	19%
Archaeology, ancient cultures and society, and cultural conservation	21	6	10	5	29%	48%	24%
Biomedicine	21	12	0	9	57%	0%	43%
Handicrafts and design	21	7	9	5	33%	43%	24%
Occupational therapy	20	2	17	1	10%	85%	5%
Physics	20	6	11	3	30%	55%	15%
Pharmaceutics and similar main fields of study, and bachelor of science in pharmacy	18	9	7	2	50%	39%	11%
Anthropology and ethnology	18	4	13	1	22%	72%	6%
Postgraduate diploma in midwifery	18	3	15	0	17%	83%	0%
Philosophy	18	1	10	7	6%	56%	39%
Tourism science and similar areas	17	8	7	2	47%	41%	12%
Agriculture, forestry and gardens	15	1	13	1	7%	87%	7%
Radiography programmes	15	2	12	1	13%	80%	7%
Biomedical laboratory science	14	3	11	0	21%	79%	0%
Gender studies	14	0	11	3	0%	79%	21%
Stage performance	14	2	10	2	14%	71%	14%

Statistics	13	4	7	2	31%	54%	15%
Postgraduate diploma in psychotherapy	13	7	4	2	54%	31%	15%
Sport and fitness sciences	13	2	9	2	15%	69%	15%
German	12	1	7	4	8%	58%	33%
Medicine programmes	12	6	6	0	50%	50%	0%
Fine arts	12	1	9	2	8%	75%	17%
Human work science	11	1	9	1	9%	82%	9%
Speech therapy	11	0	10	1	0%	91%	9%
Food sciences, nutrition and dietetics	11	3	8	0	27%	73%	0%
Archival, library and museum studies	11	3	8	0	27%	73%	0%
Journalism	10	5	4	1	50%	40%	10%
Development studies etc.	10	5	3	2	50%	30%	20%
Economic history	9	1	7	1	11%	78%	11%
Bachelor of science in biomedical laboratory science	9	2	7	0	22%	78%	0%
Cultural studies programmes	9	0	7	2	0%	78%	22%
History of ideas	8	0	6	2	0%	75%	25%
Nation studies	8	2	6	0	25%	75%	0%
Music	8	0	6	2	0%	75%	25%
Audiology	7	0	7	0	0%	100%	0%
Medical physics	7	0	5	2	0%	71%	29%
Technical translation and translation	7	1	5	1	14%	71%	14%
Occupational health	7	3	4	0	43%	57%	0%
Farming and animals	7	1	5	1	14%	71%	14%
Cognitive science	6	0	4	2	0%	67%	33%
African and Asian languages	6	1	5	0	17%	83%	0%
Linguistics	6	1	4	1	17%	67%	17%
Maritime training and similar areas	6	0	5	1	0%	83%	17%
Optometry	5	3	2	0	60%	40%	0%
Baltic and Slavic languages	5	0	4	1	0%	80%	20%
East Asian languages	5	0	3	2	0%	60%	40%
Classical languages	4	2	2	0	50%	50%	0%
Rhetoric	4	1	3	0	25%	75%	0%
Finno - Ugric languages	3	1	2	0	33%	67%	0%
Master of science in pharmacy	3	1	2	0	33%	67%	0%
Military science programmes	3	2	1	0	67%	33%	0%
Bachelor of arts in study and career guidance	3	0	3	0	0%	100%	0%
Orthopaedic technician	2	0	2	0	0%	100%	0%

The overview[23] primarily demonstrates quantity and variation, i.e., it illustrates the number of evaluations carried out in the cycle; it shows the variation in the number of programmes included in these evaluations; and it shows the variation in outcomes between these evaluations. It is difficult to see any patterns. Clearly, most of the 93 evaluation projects (approximately two thirds) resulted in all three overall judgements (very high, high or inadequate quality), though with different distribution. The remaining approximately one third of projects only received one or two

---

23. Annex 5 has an overview sorted by evaluation rounds.

of the judgements. Of these, two thirds comprise evaluation projects with fewer than ten programmes. Thus, there seems to be a difference in outcome that is connected to scope. This difference could be due to more time being allocated per programme in smaller projects, but it could also be due to fewer programmes result in fewer judgements.

The differences in scope do not seem to be a decisive factor for outcome, at least not in the 2011–2014 cycle. But the evaluations during the period had limited scope in terms of what would be judged and addressed in the reports. These factors may have facilitated the calibration for equivalence in general. The fewer programmes being assessed, most likely, the easier it is to calibrate.

The weakness of this approach was that it limited the number of aspects of quality being assessed, and the assessment texts were perceived as providing reviewees with too little feedback. Such feedback might be desirable, but it can be problematic. In a review system that strives for greater scope in both what is being assessed and how it is being assessed, the challenges for calibration of the assessments increase, especially in larger projects. In addition, feedback may be controversial if it is perceived as excessively instructive.

The latter challenges are relevant for UKÄ with the quality system that was implemented in 2017. It has a larger scope and aims to provide HEIs with detailed feedback. To handle these challenges, UKÄ introduced more steps in its assessor training, which gathers assessors from various projects for an introduction to the reviews and includes a discussion of common questions, procedures and assessment criteria. Furthermore, the ambition is for the evaluations to be instructive and to collect and disseminate good examples.

The ambition is also to summarise per review the experiences at a national level, for example via feedback conferences, and also through richer introductory reflections in the reports, which were essentially entirely lacking from the previous cycle.[24] By summarising in this way, important national experiences can be collected and distributed to HEIs and decision-makers. The format for these national snapshots is being tested in the framework for the ongoing first rounds of reviews in the new system. They will probably need to be adapted to the various components and projects.

Significance of self-criticism

Are there programmes/disciplinary domains that practice a greater degree of self-criticism than others and which consequently appear as inferior in the context of an evaluation, when in fact this is due to different approaches and academic cultures?

The overview of outcomes in the different rounds and projects above shows fairly significant variation. It seems difficult to conclude whether certain disciplinary domains in this context could have systematically fared worse due to an internal culture of self-criticism relative to other areas that have no such culture. However, this does not mean there are no differences in approach between different main fields of study and programmes, which are expressed in the individual evaluation. It is not unreasonable to assume that the paradigm of a main field of study or subject is significant in a review

---

24. However, they were included in the reviews carried out in even earlier cycles, such as 2001–2006.

and then also depending on how the review is carried out. In addition, the composition of the individual assessment panel is likely to play a very significant role.

The collegial assessment forms the framework of UKÄ's reviews, i.e., reviews are made by experts in the relevant area. This serves as the foundation of much of the legitimacy of the reviews. The assessors appointed for an evaluation are primarily nominated by the individuals who they will subsequently evaluate.[25]

Different academic cultures are thus likely to influence the placement of the bar in the different evaluations. It is both a strength and weakness of the collegial model: acceptance of being reviewed by one's colleagues, but not being challenged by others. In a way, this turns each evaluation of similar programmes into an internal matter and renders comparisons of outcomes with evaluations of programmes in other areas impossible. Even if an evaluation method is precisely applied to all evaluations, the outcome of the evaluations will therefore not represent comparable differences in quality with any certainty.

It can be added that the entire method depends on the critical examination of the assessor; development takes place in front of this individual. The challenge of reviews, regardless of who conducts them, is engaging and guiding them without also limiting them, because an assessment without criticism probably results in neither feedback nor learning.

Before introducing the new system, UKÄ has strived to determine the best forms of review through ongoing consultations with representatives of HEIs and other stakeholders, such as student organisations and labour market representatives, where they have discussed design and implementation. UKÄ also aims to strengthen the calibration of assessments in various projects and, as stated above, has therefore introduced joint training programmes for assessors from different, simultaneously occurring projects.

Context

Can the context in which a programme is reviewed affect its outcome? Can a deviation from the norm be disregarded? And could a wide array of different main fields of study in an evaluation make it more challenging for assessors to notice particular details that could be significant to the review?

One administrative condition for the implementation of comprehensive and systematic evaluations in a given time period is a subdivision, and thus categorisation, of the relevant objects.

For the sake of administrative simplicity for both HSV/UKÄ and the HEIs, the goal of each round was primarily to gather programmes in similar areas. However, it was impossible to consistently uphold this subdivision due to the varying number of programmes in different areas.

---

25. If there are no nominations or not enough of them, UKÄ must try to find experts on its own. This is usually done in consultation with the assessors who have already agreed to participate. The two completely crucial factors in this context are having the proper expertise and no conflict of interest; see more in the method annex.

The principle behind placing projects into cycles reflects an ambition for the individual projects to mainly be comprised of programmes in similar main fields of study and professional qualifications. Before each round, the HEIs were granted the opportunity to choose the context for the programmes to be reviewed.

Before each new round, a letter was sent to the HEIs' vice-chancellors with information about the upcoming evaluation projects. In the framework of each evaluation project, HEIs were also checked with, as far as possible, to find the most suitable context for the individual programme.

However, there are many main fields of study. One reason for this is that the HEIs themselves are able to define and introduce main fields of study, unlike professional qualifications, where this is governed by the Ordinance (though it is possible to introduce different focuses in addition to what is stipulated in the Ordinance). Thus, there are far too many individual main fields of study for the projects to be sorted according to these. One question is whether individual main fields of study that have been combined with others may have been disadvantaged by the context in which they were included.

Of the 93 projects included in the cycle, only 11 consisted of programmes in one single main field of study or one professional qualification. These included evaluations of the master of science in pharmacy (which included 3 programmes); bachelor of science in biomedical laboratory science (9 programmes); biomedical laboratory science (14 programmes); English (33 programmes); gender studies (14 programmes); journalism (10 programmes); rhetoric (4 programmes); bachelor of science in nursing (24 programmes); statistics (13 programmes); bachelor of arts in study and career guidance (3 programmes) and German (12 programmes).

The other 82 projects thus contained different main fields of study and/or professional qualifications. The media and design evaluation project, for example, contained 20 different main fields of study, including image production, industrial design, furniture design, and marketing and PR. Another example is the education, didactics and educational sciences project, with 19 main fields of study including action learning, dance education, educational work and adult learning.

Annex 3, which can be found on UKÄ's website, presents a list of all the different main fields of study and focuses of professional qualifications reviewed in the cycle, of which there are over 600. If engineering focuses[26] are combined with master's and bachelor's programmes in engineering, the list is reduced by about 100 titles. The purpose of the list is to demonstrate the wide variety of main fields of study which the HEIs have established.[27] It shows that just under one hundred of the main fields of study and professional qualifications have had more than four programmes, and that more than half have had one programme. This means that most of the

---

26. To manage the quantity of engineering programmes, they were sorted according to the formerly applied concept of technical field, which was used to create a structure for the evaluations of these many different programmes.

27. But perhaps the number of different main fields of study in a project need not always indicate significant variation in content; in all likelihood, the titles occasionally differ more than the content, at least, this seems feasible given the number of main fields of study (and professional qualifications).

evaluations comprised multiple main fields of study, many of which were individual.

The notion that programmes with unique titles and associated special profiles are systematically disadvantaged by their context is probably not accurate, since a large variation in evaluations tends to be the rule. However, it cannot be ruled out that individual main fields of study suffered due to the context in which they were included,[28] but it is hard to draw any conclusions about which ones based on the available data.

Interdisciplinary programmes

Has the method in the 2011–2014 evaluation system disadvantaged interdisciplinary programmes (general qualifications) through the focus on immersion in a limited area, with too little consideration for their more cross-disciplinary nature?

Several different concepts exist to describe cross-disciplinary programmes that are not easily classified as traditional disciplines, including interdisciplinary, multidisciplinary, cross-disciplinary and so on. There is a large variation in terminology and, to some extent, reflects both the degree of interdisciplinarity and the degree of integration of the various areas being combined.

This varied use of the concepts reflects disagreement in their meaning and where to differentiate between them. A previous survey of the use of these terms in evaluation contexts shows they are common but their use varies to describe more or less unique traits in different programmes. This therefore needs to be considered in an evaluation.[29]

Lack of such this type of consideration has been highlighted as an objection to the system employed in 2011–2014. And compared with the 2001–2006 cycle, the difference is clear. The review of evaluation reports carried out for the period of 2001–2005 revealed that one of these terms was mentioned in 48 out of 54 reports. The latest evaluation cycle looks quite different. Out of 93 projects, only two of the evaluation reports refer to one of these terms: the evaluations of nation studies and environmental science, respectively. In the former, the primary concept used is 'multidisciplinary'; 'interdisciplinary' is used in the latter.

It is unlikely that these are the only two evaluation reports with this discussion because they are the only projects with programmes to which these concepts may have applied. One reason why these terms only appear in two evaluations could be that the reports had limited space for such comprehensive discussions about the nature and composition of the programmes. In accordance with the template presented to the assessment panels before each evaluation, most of the reports in the cycle include a very short general write-up of how the evaluation was carried out. There was space for special additions, and these do exist, but not as a rule.

---

28. Finding individuals with adequate expertise to carry out evaluations of the various programmes has been and is of crucial importance for the reliability of the evaluations. The method annex to this report describes the recruitment process for the assessments and the distribution of degree projects for evaluation by assessors with the proper expertise.

29. Att utvärdera tvärvetenskap – reflektioner utifrån Högskoleverkets utvärderingar [Evaluating interdisciplinarity – reflections on the basis of evaluations made by the National Agency for Higher Education] 2001-2005, 2007:34 R.

Another explanation for why these terms are not used could be that the method, which focused on goal attainment, rendered them less relevant in the context. For all programmes that lead to a certain qualification (general in a main field of study or professional qualification), the same qualitative targets apply according to the relevant ordinance. Thus it would be complicated to then assert in a review that the objectives are not applicable to certain programmes. It is not obvious which objectives should be eliminated or replaced by others, and why to do so only for a certain programme.

On the other hand, it could be maintained that the data were less suited to the assessment of goal attainment for certain programmes. But the idea of equal treatment argues against this. The government bill stipulating the guidelines for the evaluations emphasised that they should be equivalent, transparent and predictable. To achieve this, the method was streamlined: everything was about goal attainment and all programmes were permitted space to demonstrate goal attainment through the same data. There was also space for special considerations. In the individual evaluation, the assessment panel was allowed to choose which qualitative targets (according to the selection principles determined by HSV/UKÄ) to include in the review, and also to decide which data would form the primary basis for assessment.

Most assessment panels, based on these conditions, have probably been able to accept assessing goal attainment based on their expertise in the relevant main field of study. They have thus avoided a discussion of any special conditions that could influence the outcome, especially negatively, and the need for special exceptions or consideration. The efforts have focused on goal attainment, not on circumstances that may have influenced it. It was important to find the proper expertise among the assessors to judge specialisations. Assessor expertise, including the subject experts who reviewed the degree projects, must reflect the programmes with specialisations that were included in each evaluation. Taking into account the different qualifications, main fields of study and specialisations that could be included in an evaluation, for the actual evaluation, the HEIs typically marked each degree project for review with its particular specialisation. The aim was to be able to sort them for subject experts, who indicated corresponding areas of expertise.

It could still be the case that the restrictions and limitations of the method meant that a certain type of programme would be disadvantaged and judged more harshly than it might have been otherwise, and thus more harshly than other programmes. On the other hand, based on the outcome, it is difficult to discern a pattern regarding what type of programme this could involve.

The outcomes for the two evaluation projects that mentioned the terms 'multidisciplinary' and 'interdisciplinary' do not offer any clear exceptions. Nation studies included eight programmes, each of which comprised its own main field of study. Six were given the grade of high quality and two were given the grade of inadequate quality. One fourth with inadequate quality is consistent with the total outcome. After their follow-ups, the inadequacies in these programmes were assessed as remedied.

Environmental science encompassed 39 programmes split into ten main fields of study, with environmental science as the major one with

25 programmes, five of which had a specialisation. The judgements were distributed as follows: three were given the grade very high quality, 28 high quality and eight inadequate quality (approx. 21 per cent). After their follow-ups, the inadequacies in all of these programmes were assessed as remedied.

One of the main fields of study in the evaluation of environmental science was environmental and health protection, with five programmes. Three of these were given the grade of inadequate quality, that is to say, more than half. The introductory text in the report describes environmental and health protection as a professionally-oriented main field of study which is therefore complicated to assess based on requirements that largely concern traditional disciplines, such as objectives for methodological expertise and an overview of the main area's methods. The next paragraph concludes, however, that inadequacies occasionally arise when it comes to a critical approach to the chosen methodology, and discusses the results obtained in the study. The discussion section often comprises simple summaries of the methods for the degree projects and results, rather than a critical discussion with conclusions.

The conclusions in the evaluation can be seen as a confirmation of the method's strength, which was in equivalency. Despite their remarks on the special character of the main field of study, the assessors conclude that in their degree projects, students must still achieve and be able to demonstrate certain knowledge and abilities in accordance with the qualitative targets. At the same time, the assessors' objections in the report must be taken seriously. They point to the weakness of the limited focus on the approach, that the judgements lack certain essential aspects of the programmes.

Taking both of these conclusions into consideration is a general challenge of evaluations, and thus also for UKÄ's ongoing and future programme evaluations. The evaluations must make sure that distinctive and multiple aspects of the programmes, in addition to goal attainment, are assessed, including through degree projects. Methods for this process should be adjusted and improved based on experience, which is now taking place within the framework of the ongoing evaluation of teacher education, which is a pilot project for method development.

Concluding thoughts

This review of the outcome for the 2011–2014 evaluation system has been based on a number of questions that emerged in the discussion about the evaluations. The questions revolve around the fact that, for various reasons, the method and implementation of the evaluations may have disadvantaged certain types of programmes. The discussion in the report touched upon the fact that it is possible to draw certain conclusions about these questions and assertions based on the actual outcome of the evaluations, for example, based on any patterns in the outcome.

The simple summaries made here do not clearly demonstrate any such connections. But that does not mean the questions are uninteresting. One point here is that they are general in nature and can therefore be considered challenges to take note of in all evaluations. There is always cause to take note of opinions and to evaluate the method of review. The review of the questions in the report also mentions changes that UKÄ made in the

evaluations in the new system for quality assurance, which applies from 2017. Among other things, this includes the use of new terminology; a focus on assessing the whole rather than parts; and a shift away from a purely results-oriented approach and towards awareness of the fact that conditions and processes, as well as results, should be considered. These changes aim to capture a broader spectrum of quality in the programmes than the 2011–2014 cycle.

Taken together, the changes involve not only a clear shift in focus, but also in ambition. The system for evaluations that applied during the 2011–2014 cycle strived for equivalence, transparency and predictability.[30] This meant that all programmes would be treated equally, with a focus on goal attainment as demonstrated by the data. Thus the method was somewhat biased, partly towards a data set, the degree projects, which demonstrated the outcomes most clearly, and partly towards a way of reporting the assessments, as a form of quantitatively based conclusions.

Undoubtedly, this bias contained a clarity, which touched upon the three key words to varying degrees. Transparency was found in what would be assessed and deemed decisive. On the other hand, equivalence and comparability applied especially to the assessments within the individual evaluation. This can be seen in the summaries here, which demonstrate the distribution of outcomes between the evaluations. Despite attempts for methodological stringency, compliance with the key words in the bill was difficult to achieve in the comparison between different evaluations. Even if the method strived for equivalence in all assessments, it is not possible to ignore factors that influence individual evaluations, such as discussions carried out by the assessors and where they agree to place the bar in their assessments.

The new system for quality assurance, which applies from 2017, de-emphasises the importance of transparency and equivalence. This means that the focus on assessing parts has been dropped in favour of an emphasis on the whole, where established principles for weighing the data are no longer prioritised. And instead of equivalence and equal treatment, consideration for distinctive qualities and profiles is emphasised, where the implementation of assessments should be able to be adapted based on such special conditions. However, UKÄ is also striving for better coordination between the different reviews than before. This can be seen, among other places, in the ambition to arrange common project-wide assessment trainings and checks.

Overall, major changes have been implemented, where the entire design of the new system for quality assurance, including how the evaluations are carried out, can be seen as a reaction to the specialisation and ambitions of the previous system. Time will reveal whether the changes will impact the questions used to for the reviews or if different questions will be relevant in the future. The same is true for how the outcome will be viewed, and if it will be considered better suited to assessing and improving programme quality. UKÄ intends to continue tracking, following up and analysing the method, outcomes and effects.

---

30. Government bill 2009/10:139 En akademi i tiden - ökad frihet för universitet och högskolor [Academia for this day and age - greater freedom for universities and other higher education institutions].

Annex 1. The 2011-2014 method in short

Results and autonomy

The focus on assessing results arose in reaction to previous evaluations by UKÄ's predecessor, HSV, being focused on the conditions and processes of the programmes.[31] Another important reason was the autonomy reform.[32] According to the Government at that time, the implications of increased self-determination for HEIs also entailed greater responsibility for independent quality-assurance of the conditions and processes of the programmes. The path to goal attainment was determined locally at the HEIs, while results were checked nationally.

Main principles

In practice, the foundation for assessing goal attainment in the system consisted of a selection of degree projects, the description of goal attainment in the self-evaluation and, to some extent, discussions and interview responses. The crucial point was how the degree projects would be reviewed. In the context of evaluations, systematically conducting such reviews to assess the outcome, i.e. goal attainment, was perhaps a ground-breaking element of the system. It was facilitated through the clear government assignment.

Measuring outcomes was not something for which only the Government was looking for; rather, it was a generally requested aspect and focus for the new higher education evaluation system.

Other key words for the development of the system were equivalence, transparency and predictability. These words were also highlighted in the government bill.

Object under evaluation: programme – qualification – level – course – main field of study

The Government assignment was to evaluate programmes that could lead to the award of a first- or second-cycle qualification. The third-cycle level was not included.

The concept of a 'programme' in this context was practical but vague, especially given that the Swedish higher education system is based primarily on courses and not on centrally regulated programmes. The principle entails that courses at the first- and second-cycle levels can be combined into a

---

31. Also see the report Utbildningsutvärderingarnas effekter [Effects of Evaluations]

32. Govt. bill 2009/10:149 En akademi i tiden - ökad frihet för universitet och högskolor [Academia for this day and age - greater freedom for universities and other higher education institutions]

programme of study, but students may either follow a programme or take freestanding courses.

A programme of study or a combination of freestanding courses may thus form the basis for a qualification at the first- or second-cycle level. What the qualifications share is the existence of a main field of study in which the student must have completed a certain number of credits, which is regulated.[33]

Within the main field of study for the qualification, the student must thus have achieved the Government-stipulated, nationally applicable qualitative targets for the qualification, which are regulated in the Qualification Ordinance.

General qualifications and qualifications in the fine, applied and performing arts in different main fields of study are established by the HEI with the authority to do so.[34] Sometimes a main area is unique to a certain qualification level and a certain HEI; sometimes it is found at most HEIs. A programme to be evaluated could be classified as a main field of study for a qualification. Existing main fields of study, in which a qualification may be obtained, were the starting point for what would be surveyed in preparation for the implementation of the evaluations.

In addition, a number of professional qualifications are regulated in the Qualification Ordinance. Because professional qualifications are named in the Qualification Ordinance, they are relatively easy to define and survey, even if the government assignment was unclear in how to handle different specialisations within them.

A regulated difference between general qualifications and professional qualifications is that the Qualification Ordinance stipulates specific goals for each professional qualification. Even so, it is not uncommon for a student pursuing a professional qualification to also pursue a general qualification in a main field of study corresponding to the relevant professional qualification, which is called a double degree. If a single degree project is used for both qualifications, one consequence for the evaluations is that goal attainment must be assessed based on two different qualifications.

Method development

Review of degree projects

Statistical model

A model for reviewing degree projects was needed, and principles for selection and assessment needed to be developed and determined. The government bill refers to two reports that deal with the review of degree projects, the study En akademisk fråga – en eso-rapport om ranking av uppsatser [An Academic Question – ESO Report on the Ranking of Theses] (Ds 1999:65), and a report from Lund University, Direkt kvalitetssäkring – Bedömning av utbildningens resultat och relevans [Direct Quality

---

33. Higher Education Ordinance, Annex 2 Qualification Ordinance
34. See UKÄ's website.

Assurance – Assessment of Educational Outcomes and Relevance] (2009:255). The reports are referenced to validate the argument that degree projects are relevant for the assessment of programme outcomes, because they summarise much of the knowledge and skills which higher education should provide. However, these reports do not answer how to make selections for national reviews weighing certainty of the assessments on the one hand and costs in monetary and assessor effort on the other. The scope of work that went into the review in these studies was far too limited, and the approach was not directly transferrable for the task at hand. Statistics Sweden was consulted, whereupon it was concluded that a model which asserts with statistical certainty the quality of each programme through the division of quality of the degree projects was unrealistic. A very large number of projects would need to be reviewed for the bigger programmes, which would be costly and labour intensive for the assessors.

Nor were such variations as random selections or small, strategic selections deemed suitable, given the variation in programme size and the potential significance of the selection. Thus, the idea for a model that showed the distribution of the quality of degree projects was discarded.

Instead, HSV (UKÄ from 2012) designed a model based on random samples. The samples were tested to ensure they were of such a size that they could ensure with some certainty that the selection included projects of varying levels of quality from the population. Based on these, it would be possible to make reasonable assessments of the presence of variations in quality. Depending on the number of degree projects in the programmes, the number of projects in the selection varied according to the model from 5 to 26.

Fewer than five and comparability

A minimum selection of five degree projects meant that programmes with fewer than five projects could not be assessed on this basis. To avoid excluding data for more programmes than necessary, projects examined in the most recent three-year period were counted to reach the limit. After initial attempts during the first round of evaluations, the goal to also evaluate programmes with fewer degree projects was abandoned.[35] An important reason was the challenge of maintaining comparability when using different data. Comparability was one of the key words for the system, which was emphasised in the government bill and assignment.

Qualitative targets as aspects of quality

How to assess goal attainment in degree projects was another question. In the bill, the Government at that time noted that experience from the study of principles for reviewing the degree projects at Lund University showed an opportunity for identifying quality aspects shared by higher education programmes that were significantly different in nature. The assignment for UKÄ thus included the formulation of such quality aspects. Proposals were also developed, including a connection to research/artistic basis, subjects/

---

35. Högskoleverkets system för kvalitetsutvärdering [National Agency for Higher Education's System for Quality Evaluation] 2011–2014, Report 2012:15 R, p. 9–10.

problem formulation, method, implementation, documentation and problem-solving.

Selection of qualitative targets

The number and content of qualitative targets varies for different qualifications. A review of all objectives for each qualification would create enormous differences in the scope of the evaluations. To reduce these differences, a general selection principle was applied for a certain number of objectives to be judged per evaluation. The principle involved the requirement to review a given range of objectives per form of knowledge. Thus, a minimum number of qualitative targets per evaluation was established, while a variation range was defined to account for the unique nature of each programme as seen in the differences between the objectives for different qualifications. This also reduced the work load for both the HEI and for the assessors. The selection of qualitative targets, which would take place before each individual evaluation, was done by the panel of assessors that would subsequently conduct the evaluation.

A total of 263 individual objectives were selected, of which 228 applied to professional qualifications. There is no compilation of all goals in the Qualification Ordinance, rendering it impossible to know the percentage of goals from the Qualification Ordinance included in the reviews.

Goal attainment in degree projects, not a review of grades

Another important idea was that the degree project assessments were not to be a form of review of the grades given at the HEIs. Therefore, it was not relevant to assess the quality of each project as a whole; rather, what was assessed was goal attainment in the projects for the selection of qualitative targets and pursuant to the principles determined by the assessment panel for each evaluation. Goal attainment was compiled per objective for the selection of degree projects and for each programme evaluated.

Assessment and importance of strengths and weaknesses in individual projects

The government assignment stipulated that 'individual' strong or weak performance should not form the basis of the assessment. This entailed a general principle that more than individual projects indicating inadequate goal attainment was required for a judgement of 'inadequate goal attainment'. The limit for 'more than individual projects' was set so that the share of degree projects with inadequate goal attainment for one objective comprised more than 30-40 per cent of projects in the selection. The range was considered necessary, partly because the selections varied in size and, thus, so did the impact of the assessments of individual projects; and partly because the assessed level of goal attainment for each individual project could differ for the same judgement, because 'there are weak and strong lowest, medium and highest ratings'. The idea was that this variation should be able to be observed via the range to some extent. Using double readers for degree projects was not possible for reasons of both time and costs; rather,

efforts were made to achieve the principle that assessments of individual projects would not be conclusive.

Degree projects and other data

The government assignment included the following: 'Together with educational outcomes reported in the self-evaluations, the degree projects should comprise the main portion of the overall judgement, wherein most cases, the degree projects are given the most weight. For certain programmes, there are reasons to increase the weight of the reporting of outcomes in the self-evaluation. For example, in the event that the degree projects do not reflect important objectives for programmes...'[36]

This principle also came to guide the evaluations. Guidelines for the weighting of data in accordance with the assignment were also developed.[37]When the assessment of the degree projects clearly indicated the outcome, according to the assessors, it was difficult to change the conclusion based on other, more process-oriented data. Therefore, the reports on goal attainment were occasionally short, when goal attainment could be assessed through the degree projects and the projects provided a clear indication.

Assessment panel, subject experts and review of degree projects

The HEIs nominated assessors for each evaluation. HSV/UKÄ appointed the panel. In addition to the assessors included in the panel that summarised the outcome and formulated reports, several evaluations included subject experts tasked with assessing goal attainment in the degree projects. The subject experts' competence needed to reflect the programmes with special focuses that were included in each evaluation. Adequate and legitimate assessment was essential. Taking into account the different qualifications, main fields of study and focuses that could be included in an evaluation, the HEIs typically got to define the projects to be reviewed with their respective focuses so these could be sorted for subject experts with corresponding areas of expertise.

Before reviewing the degree projects, the assessment panel members and subject experts who would contribute reviews of the relevant projects held meetings to discuss and calibrate principles for assessment. These meetings ordinarily involved several example degree projects that would not be included in the selection of projects to be reviewed in the evaluation.

The self-evaluation

In the government assignment, the self-evaluation was allocated two purposes. One purpose involves the reporting of conditions with clear significance for educational outcomes, as noted previously in this report. The other purpose, which was designated most important, is to report on

---

36. The annex to the government assignment on 8 July 2010 for the National Agency for Higher Education to develop a quality evaluation system.

37. Government bill 2009/10:139 Fokus på kunskap – kvalitet i den högre utbildningen [Focus on knowledge – quality in higher education], Annex to the government assignment to the National Agency for Higher Education on 8 July 2010, Högskoleverkets system för kvalitetsutvärdering 2011–2014 [National Agency for Higher Education's system for quality evaluations 2011–2014], Report 2012:15 R.

outcomes that are not shown through degree projects, especially when considering specific requirements for different qualifications. The first purpose still remains in HSV's system document. The second purpose has been slightly revised: the text indicates that all objectives addressed in the evaluation must be included in the self-evaluation. The HEIs must show, analyse and evaluate the outcomes in relationship to the qualitative targets.

One possibility that was discussed was to request exam questions and answers. Doing this systematically would be burdensom for both HEIs and assessors. Another suggestion was that similar data could be used to exemplify goal attainment instead. Otherwise, it was difficult to provide specific examples of how goal attainment should be reported without simultaneously becoming standardised in a system, where one main point was that the paths to goal attainment were decided by the programme designers. The task of describing goal attainment was thereby passed largely to HEIs, with the assumption that the burden of proof was on them. They issued a qualification which presupposed that passing students had achieved the objectives.

Working life and student perspective

One aspect emphasised in the assignment and in the government bill was the significance of the working life perspective in the evaluations. Pursuant to the stipulations in the bill, consultation was arranged with labour market representatives before each evaluation.

Labour market representatives and students were also included on the assessment panels, in accordance with the assignment. As with the consultations mentioned above, an overall evaluation of this collaboration is lacking. Generally speaking, their collaboration was more meaningful for the discussions prior to evaluations and the summaries than for assessing whether the data sets for a programme indicated goal attainment.

Structure of the evaluation cycle

One administrative condition for the implementation of comprehensive and systematic evaluations in a given time period is a subdivision, and thus categorisation, of the relevant objects.

The division of main fields of study into clusters is an important task for the main fields of study and professional qualifications to be reviewed in contexts that make the assessment as fair as possible and lend it legitimacy. One question is whether each individual main field of study, or specific focuses of each one, should be reviewed individually for an assessment that is as adequate as possible, in which individual characteristics are taken into consideration. Such a strict division is hardly feasible within the framework for a national valuation system. In addition, there may be advantages to gathering programmes within a main field of study and similar areas, in the framework of a review, for a greater exchange of experts, perspectives and experiences. A broader grasp could expand opportunities for different angles of approach and perspectives within and connected to individual main fields of study to spread experiences.

To provide the best conditions for considering the multitude of main fields of study and their focuses, UKÄ, formerly the National Agency for Higher Education (HSV), has divided them into clusters in several steps. Before the evaluation cycle, the HEIs' main fields of study were mapped out, with a focus on relevant qualification levels. Based on this mapping process, a proposal was made for how to structure the division of the evaluations, which the HEIs were allowed to comment on. After these were considered, the structure was posted on the government agency's website. This design meant a division into six rounds which have consisted of about 10–15 evaluation projects. For the sake of administrative simplicity for both UKÄ and the HEIs, the goal was for each round to primarily bring together programmes within one disciplinary research domain. However, it was impossible to consistently uphold this subdivision due to the varying number of programmes in different areas. Before each new round, a letter from the vice chancellor was sent to the HEIs to sort existing programmes. In the framework of each cluster or evaluation project, each HEI was consulted to, as far as possible, find the most suitable context for the individual programme.

In actuality, one starting point has been to divide the relevant programmes by name: main areas[38] for general qualifications (such as the main fields of study business administration, gender studies, medicine) and title for professional qualification (master of science in medicine, bachelor of science in engineering).

It was not possible to address each individual main field of study or individual professional qualification separately, because the number of main fields of study and professional qualifications is far too high. In total, over 500 main fields of study and professional qualifications were reviewed during the period. It was necessary to combine similar main fields of study and professional qualifications into clusters (see annexes, available via UKÄ's website). Based on this principle for division, the evaluations were carried out in 93 projects, usually named after the largest main field of study or the largest professional qualification included, for example the evaluation of business administration and similar main fields of study.[39]

---

38. Corresponding with what were called subjects prior to 2007.

39. The plan for the cycle indicates 88 clusters. Three of them were divided into additional clusters, of which 93 are presented here. The projects began in six different rounds, spread out over the cycle, which formally took place from January 2012 through October 2015.

The Swedish Higher Education Authority (Universitetskanslersämbetet – UKÄ) is to contribute to strengthening Swedish higher education and Sweden as a knowledge society. We review the quality of higher education programmes; we analyse and follow-up trends within higher education and we monitor the rights of students.

www.uka.se