10th European Quality Assurance Forum

19-21 November 2015

Quality Assurance Agency and UCL Institute of Education

London, UK

Taking stock and looking forward

Paper presented during EQAF 2015

Author(s)

Name: Jon Haakstad

Position: Senior adviser

Organisation: Norwegian Agency for Quality Assurance in Education (NOKUT)

Country: Norway

Short bio:

Address	Stasjonsveien 47A, 0771 OSLO, Norway
Former positions	1974 – 1992: Associate professor in English, Tromsø College of Education 1992 – 1994: Rector, Tromsø College of Education 1994 – 1996: Assistant director, Tromsø University College 1996 – 1998: Office manager, Norwegian Teacher Training Council 1998 – 2002: Senior adviser, Network Norway Council 2003 – 2008: Director of audits and evaluations, NOKUT 2008 – 2011: Director of analysis, NOKUT
Writings: and	*History of Britain for Students of English* (1998, textbook) Articles, conference papers and public reports on educational quality quality assurance in higher education
Intern. activities:	External expert in several evaluations of higher education (Denmark, Ireland, Austria and Switzerland) External expert (chair) in ENQA evaluations of QA agencies (Finland, Austria, Spain, Germany and the EUA Frequent presenter at EAIR and EQAF conferences Member of AQA's (Austria) Scientific Steering Group (2008 – 2013)

Proposal

Title: What do the indicators tell?

Abstract:
Input factors have always been an important element in the evaluation of education. Another approach is to highlight outcomes, and yet a third to listen to the students' assessment of

course and programme quality. Ideally, correspondences between these three perspectives should make up a logical picture; performance indicators on factors like intake, input and outcome, plus the students' appreciation of quality, should present a consistent body of information that allows us to make reasonable assessments of the quality of provision. But to what extent do we actually find these correspondences and this consistency in the data? In other words: Can we rely on the indicators that we refer to so often in our quality discussions? A study of certain key data for all Norwegian BA and MA programmes in biology, political science and nursing suggests that performance indicators tell us much about performance but may indicate less about quality.

Text of paper:

What do the indicators tell?
Three dives into a large pool

Topic
Assessment of input factors has always been an important element in the evaluation of education. Another approach is to take one's point of departure in outcomes, which seems no less important with the Qualifications Framework and its set aims for learning outcomes. A third approach is to listen to the students' assessment of course and programme quality. In Norway we now have fairly good data for both inputs and outcomes, while the new National Student Survey provides data on students' assessments at programme level across the sector. To what extent does the available information allow us to see correlations between these three perspectives on educational quality?

Rationale
It is a commonplace that we cannot really have an exact description, let alone exact assessments, of educational quality. The theme is too large; the concept of quality is too complex; too much of the educational process is hidden from our view. Still, quality is usually referred to as the reason for reform efforts, like for instance the current process of institutional mergers in the Norwegian sector. We then make use of *indicators*, although everyone knows that they are simplifications. It is therefore of some importance to have an understanding of the indicators' information value.

Interpreting indicators is fraught with problems. An institution's total input is so much more than what a few select indicators can tell; the students, when assessing their programmes, can hardly have the full picture of what is important and relevant and what the right level is; formal outcomes are typically surrounded by uncertainties concerning the use of the grading scale[1], while the supposed relationship between inputs and outcomes is disturbed by variations in the students' abilities, their study effort and their previous learning. This prompts some interesting questions:

- To what extent should we trust formal results as expressions of the students' learning status[2] after their completion of a programme?

---

[1] A recent study in Norway gives cause for alarm: Strøm & al: *Quality and the use of the grading scale in Norwegian higher education* (My translation of a title in Norwegian, JH). The study indicates vast differences in the use of the grading scale.

[2] It is the students' learning *status* that is assessed at the end of a programme. It is much harder to make an assessment of their learning *outcome from the programme,* as this is mixed with knowledge and competencies acquired from other sources.

- What correlations – or not – can we detect between obvious intake and input indicators on the one hand, and outcomes on the other?

Choice of sample and indicators

This study relates to the Norwegian HE sector and the choice of disciplines has been made from considerations of type, size and the quality of data. The choice is otherwise random. All programmes at MA and BA level in biology, political science and nursing where good data are available are included. The choice of indicators has been determined by what factors are most commonly referred to in discussions of quality. The following are included:

- **Intake**
  - *Students' score from secondary school[3]*
- **Input**
  - *Competence of discipline community in terms of percentage with 'first position competence'[4]*
  - *Discipline community's total publication points in national research database[5]*
  - *Students' assessment of teaching quality [6]*
  - *Total time (weekly average) that students spend on their studies*
- **Outcome**
  - *Percentage of students achieving grade A or B[7]*
  - *Average number of ECTS credits achieved by students annually.*

The students' score from secondary school is not an input factor in the traditional sense but is still expected to influence both learning process and outcomes. 'Intake quality' obviously includes more than this, but it may still be the best single indicator of ability and aptitude.
As for the two indicators that describe the academic strength of the providing discipline community, both are central in quality discussions: accreditation processes often turn around the question of the teachers' formal (research!) competence, while publication points illustrate the size and the research activity of the discipline community, both crucial elements in the current demand for larger, more 'robust' academic units. Quality has to some extent been equated with size and research.
Unfortunately, there is no indicator that yields systematic data on the teachers' didactic competence. Nor do we have national figures on the volumes of teaching invested in the programmes. And since national programme evaluations are not carried out except for select control purposes we do not have sector-wide qualitative assessments of teaching and learning processes either. What we do have is the students' assessment of teaching and academic counselling, as these appear in the National Student Survey. The Student Survey also contains information from students concerning the time they spend on their studies.
The two outcome indicators have been picked because they tell something about (a) what percentage of the students 'do well' and (b) the effectiveness of programmes in bringing their students to completing their degrees.

Method – and some reservations

---

[3] Source: The Norwegian Universities and Colleges Admission Service.

[4] 'First position competence' is a term used in Norway for academics with doctorates or equal qualifications (professors and associate professors). Source: The Norwegian Social Science Data Services

[5] Source: The Norwegian Social Science Data Services

[6] Source for this and the next ('total time') indicator: NOKUT's National Student Survey

[7] Grading scale from A to F (Fail). Source for both outcome indicators: The Norwegian Social Science Data Services.

The scores of the individual institutions on the series of input and student assessment indicators are compared with scores for outcomes in search of correspondences. Results from the two chosen indicators for outcomes are also compared with each other. Since the way of expressing scores is different from one indicator to another, the institutions' scores are ranked for each indicator to provide a way of comparing them.

Some reservations have to be made with reference to limitations in the data:

- Based on the most recent statistics, scores for intake and outcomes will refer to different year groups in the same programme.
- The scores may refer to more than one programme, as they include for instance all biology programmes at the individual institution.
- As there are no available data at discipline level for competence profiles and research, scores for these indicators will refer to the relevant department or faculty.

Although these limitations may cause uncertainties about the validity of comparisons, their importance should not be exaggerated either. One must assume that a considerable degree of stability prevails in intake and input factors at each individual institution; likewise, the characteristics of the academic department/faculty community will reflect on the narrower discipline community.

Findings
Space does not allow a full presentation of data and analysis for all three studied disciplines. **Biology** will have to serve as an illustration. Main findings in **political science** and **nursing** are then compared with biology. Finally, a few reflections are made on the background of the findings.

A) Biology

Indicator scores

Insti-tution	Intake	Comp[8]	Publ[9]	Teach	Time	A+B	Cred	No. of stud
A	42,9	94,3	300	3,6	35,0	43,4	37,3	848
B	47,8	97,2	571	3,2	38,6	39,4	49,4	431
C	41,2	81,3	161	3,9	33	40,8	49,4	98
D	42,1	99,6	582	3,3	37,6	41,8	43,5	620
E	44,4	96,4	29	3,5	28,5	43,9	42,6	168
F	41,3	95,5	1020	3,3	38,5	38,6	42,0	632
G	42,7	91,4	228	0[10]	0	43,5	41,3	2
H	43,2	89,1	151	0	0	41,9	50,2	212
I	40,7	59,9	19	0	0	39,7	47,2	26
J	35,2	57,9	75	3,9	44,5	53,0	46,8	73
K	38,8	51,6	13	0	0	55,2	47,5	14
L	37,8	66,3	9	0	0	34,4	48,4	175
M	38,5	88,1	65	3,4	37,4	44,0	42,8	246
N	31,6	43,4	5	3,1	30,3	24,9	41,5	86

---

[8] Figures for relevant department

[9] Figures for relevant faculty

[10] Zeroes indicate that the number of responses was too low for publishing in the National Student Survey.

Indicators

Intake: Average school leaving points; **Comp (Competence):** Percentage of faculty with first position competence; **Publ (Research publication):** Faculty's registered publishing points in national database; **Teach:** Quality of teaching, assessed by students in National Student Survey; **Time:** Average weekly hours spent on study work, given in National Student Survey; **A + B:** Percentage of students with grade A or B; **Cred (Credits):** Average credits annually per student.

Ranking (1 – 14) by indicators:

Institution	Intake	Comp	Publ	Teach	Time	A+B	Credits
A	4	5	4	2	6	6	14
B	1	2	3	7	2	11	2
C	8	9	6			9	2
D	6	1	2	5	4	8	8
E	2	3	10	3	8	4	10
F	7	4	1	5	3	12	11
G	5	6	5			5	13
H	3	7	7			7	1
I	9	11	11			10	6
J	13	12	8	1	1	2	7
K	10	13	12			1	5
L	12	10	13			13	4
M	11	8	9	4	5	3	9
N	14	14	14	8	7	14	12

Analysis:

- *Outcome indicators compared*
  The six institutions with the highest credits production are ranked as nos. 1, 7, 9, 10, 11 and 13 (of 13 altogether) on 'percentage with A or B'. The best three are nos. 7, 9 and 11. The six highest-scoring on 'A or B' are nos. 5, 7, 9, 10, 13 and 14 on credits production. The three highest are nos. 5, 7 and 9. The distributions seem arbitrary, with no visible correspondence between the two outcome indicators.

- *Intake – outcome*
  The six institutions with the highest average intake level rank as nos. 4, 5, 6, 7, 8 and 11 on 'A or B', and as nos. 1, 2, 8, 10, 13 and 14 on credit production.
  The six institutions with the lowest average intake level (8 – 13) are ranked as nos. 1, 2, 3, 9, 12 and 13 on 'A or B', and as nos. 4, 5, 6, 7, 9 and 12 on credit production. Overall, the bottom six score slightly better than the top six. The distributions seem arbitrary, with no visible correspondence between the intake and outcome indicators. This is particularly remarkable as the differences in intake levels are quite significant.

- *Academic community – outcome[11]*
  The top six institutions on 'competence' are nos. 2, 5, 7, 8, 11 and 12 on 'outcome'
  The top three are nos. 5, 7 and 8. This indicates arbitrary distribution.
  The bottom six on 'competence' are nos. 1, 3, 4, 8, 10 and 13 on 'outcome'. The bottom three are nos. 1, 3 and 13. Again this indicates arbitrary distribution, but the bottom six score slightly better than the top six.
  The top six on 'publishing' are nos. 2, 4, 5, 8, 11 and 12 on outcome. The top three

---

[11] For reasons of simplification the two outcome indicators are ranked as combined, and then compared with 'competence' and 'publishing' respectively.

are nos. 5, 8 and 12.

The bottom six on 'publishing' are nos. 1, 5, 7, 8, 10 and 13 on outcome. The bottom three are nos. 1, 10 and 13.

The top half rank slightly (insignificantly) better that the bottom half on 'publishing', whereas the situation is the opposite for 'competence'. The solid impression is one of arbitrary distribution.

- *Students' assessment of teaching quality compared with other indicators*
  Only eight of the 13 institutions have scores from the National Student Survey. Another 'problem' is the fact that scores in the Survey are generally so even that ranking them seems rather pointless. Instead, we select the few institutions that distinguish themselves positively or negatively on 'teaching' to see if their scores correspond with scores on other indicators:
  - o The top three on 'teaching' also score significantly better than the average on 'A or B', whereas this correspondence is lacking with 'credits'.
  - o The weakest institution on 'teaching' is also weakest and second weakest on the two outcome indicators; it is also weakest on 'intake', 'competence' and 'publishing', so here the correspondence is nearly total. But this institution seems to stand alone as the 'black sheep 'of the sample. If we move to the second weakest institution we get a contradictory picture on outcome (rankings 2 and 10 on the two indicators) and negative correspondence on 'intake', 'competence' and 'publishing' (rankings 1, 2 and 3!).

    It may seem as if absolute top or bottom assessment by the students for teaching quality is reflected in outcomes, but the picture is otherwise very contradictory, and no clear correspondences are visible.

- *Students' own assessment of time spent on studies compared with other indicators*
  One institution stands out with a particularly high figure for time spent on studies. It has a very low intake level and scores low on 'competence', whereas students give it a high score for teaching. Outcome scores are quite good.
  Two institutions stand out with particularly low figures for time spent. One of them has a medium score on outcome, the other the lowest score of all. The one with a medium score achieves high ranking on 'intake' and 'competence', whereas the one with the lowest score is also ranked at the bottom for 'intake' and 'competence'.
  For institutions with top and bottom scores in the students' assessment of teaching quality and time spent, there seems to be some positive correspondence with formal outcomes.

B) Political science and nursing - compared with biology

For these two disciplines the analysis followed the same path, but only main findings and tendencies are described here. All data are available in an appendix.

Political science
There is the same lack of correspondence between the two outcome indicators as we saw for biology; in several cases the discrepancy is 'dramatically' wide.

The intake level does not seem to have any influence on the number of students who 'do well' (i.e. get the grade A or B), while there is a weak correspondence between intake level and the number of credits achieved.

It is not possible to see any clear correspondence between 'competence' and 'publishing' on the one hand and the outcomes indicators on the other.

The top four institutions in the students' assessment of teaching quality do a little better than the average on 'A or B', but a little poorer on 'credits'. For the bottom four on teaching quality the correspondence is negative: they do slightly better than the average.

Numbers of weekly hours spent on studies are on the whole very even, so they would not be expected to result in systematic differences in outcomes. But there are two very clear exceptions where students register much fewer hours. These two institutions show a wide spread in terms of outcomes, with strikingly high scores on 'A or B' (ranked 1 and 2!) and rather weak scores on 'credits'.

Nursing

There is weak correspondence between the two outcome indicators, except for one institution that stands out as no. one on both of them. The six highest ranking institutions on 'A+B' are nos. 1, 18, 7, 16, 14 and 3 (of 18) on credits, and the discrepancies are equally big the other way.

Nor do we see high intake levels reflected in outcomes. The six highest ranking institutions are ranked as nos. 14, 1, 5, 2, 10 and 18 on 'A+B', and as nos. 6, 1, 14, 18, 15 and 11 on credits. If anything, we see a random distribution for 'A+B' and negative correspondence for credits. This is all the more remarkable since the differences in average intake level are considerable.

The institutions are very different in terms of size, institutional type and research publication. The same goes for 'competence', where the percentage of staff with first position competence ranges from 22 to 52. Still, it is difficult to see how high competence and heavy research affect outcomes in a positive way. The six highest on competence have an average ranking of 8.5 on 'A+B' and 14 (!) on credits. The picture is much the same for 'publication': 8.3 on 'A+B' and 12 on credits. The more 'robust' institutions seem to have problems with their students' progression.

The average ranking of the six institutions that score best in the students' assessment of teaching quality is 10 on both 'A+B' and credits – a slightly negative correspondence. Nor do we find that the students are more satisfied with teaching quality in 'robust' institutions. A curious finding is the fact that the four 'best' institutions on teaching quality are ranked on intake level as nos. 13, 14, 16 and 17!

Concluding reflection

The most obvious thing that springs to mind is the almost total lack of correspondences between entrance/input indicators and formal outcomes. This is so both in a factor-by-factor analysis and when all input and both outcome factors are aggregated. Nor does the students' appreciation of teaching quality correspond with outcomes, except perhaps for the very top and bottom ranked institutions. Instances of correspondence may occasionally be spotted *among* the input factors, but these are generally not reflected in the outcomes.

The lack of meaningful coherence in outcomes strongly indicates these as the elephant rogue: And when outcomes cannot be trusted, all other comparisons exist on slippery ground. But this does not explain everything: it is worth keeping in mind that although

inconsistent use of the grading scale is a main suspect here, the figures for ECTS are absolutes, and they correspond no better with inputs than the 'A+B' indicator.

So, what do the indicators tell? According to this study at least, not much. Or rather, they tell exactly what they tell: that some institutions have more professors, that some do more research, that some programmes have better recruitment and that students value programmes differently. These features may be important qualities in themselves. We may want higher education teachers to be experienced researchers, like we may want discipline communities to be 'robust'. But the performance indicators are more statements than indicators, as they do not seem to indicate much beyond themselves, like for example about educational quality.

References:

- Strøm & al: *Quality and the use of the grading scale in Norwegian higher education* (2014) (My translation of a title in Norwegian, JH)
- The Norwegian Universities and Colleges Admission Service.
- The Norwegian Social Science Data Services
- The Norwegian National Student Survey (NOKUT, 2015)

Appendix

Biology

Indicator scores

Insti-tution	Intake	Comp[12]	Publ[13]	Teach	Time	A+B	Cred	No. of stud
A	42,9	94,3	300	3,6	35,0	43,4	37,3	848
B	47,8	97,2	571	3,2	38,6	39,4	49,4	431
C	41,2	81,3	161	3,9	33	40,8	49,4	98
D	42,1	99,6	582	3,3	37,6	41,8	43,5	620
E	44,4	96,4	29	3,5	28,5	43,9	42,6	168
F	41,3	95,5	1020	3,3	38,5	38,6	42,0	632
G	42,7	91,4	228	0[14]	0	43,5	41,3	2
H	43,2	89,1	151	0	0	41,9	50,2	212
I	40,7	59,9	19	0	0	39,7	47,2	26
J	35,2	57,9	75	3,9	44,5	53,0	46,8	73
K	38,8	51,6	13	0	0	55,2	47,5	14
L	37,8	66,3	9	0	0	34,4	48,4	175
M	38,5	88,1	65	3,4	37,4	44,0	42,8	246
N	31,6	43,4	5	3,1	30,3	24,9	41,5	86

Indicators

Intake: Average school leaving points; **Comp (Competence):** Percentage of faculty with first position competence; **Publ (Research publication):** Faculty's registered publishing points in national database; **Teach:** Quality of teaching, assessed by students in National Student Survey; **Time:** Average weekly hours spent on study work, given in National Student Survey; **A + B:** Percentage of students with grade A or B; **Cred (Credits):** Average credits annually per student.

Ranking (1 – 14) by indicators:

Institution	Intake	Comp	Publ	Teach	Time	A+B	Credits
A	4	5	4	2	6	6	14
B	1	2	3	7	2	11	2
C	8	9	6			9	2
D	6	1	2	5	4	8	8
E	2	3	10	3	8	4	10
F	7	4	1	5	3	12	11
G	5	6	5			5	13
H	3	7	7			7	1
I	9	11	11			10	6
J	13	12	8	1	1	2	7
K	10	13	12			1	5
L	12	10	13			13	4
M	11	8	9	4	5	3	9
N	14	14	14	8	7	14	12

---

[12] Figures for relevant department
[13] Figures for relevant faculty
[14] Zeroes indicate that the number of responses was too low for publishing in the National Student Survey.

Political science: Indicator scores

Institution	Intake	Comp	Publ	Teach	Time	A+B	Credits	No. of stud.
A	44,9	87,3	420	3,1	29,8	46,1	45,1	330
B	40,7	83,9	74	3,7	29,5	42,5	45,6	440
C	42,7	97,2	275	3,4	30,2	37,0	42,5	925
D	38,7	81,4	56	3,2	31,8	34,1	37,9	201
E	47,9	96,4	538	3,2	31,8	45,8	40,1	1870
F	38,2	66,3	141	2,9	23,9	51,5	42,5	407
G	36,3	72,9	472[15]	3,0	29,6	33,7	47,1	462
H	41,4	74,0	40			43,8	40,1	23
I	37,5	74,0	40			35,9	37,1	293
J	36,5	61,2	17	3,6	10,8[16]	55,4	39,7	222
K	36,7	73,1	35			36,0	48,0	226
L	41,9	66,4	148			39,5	52,8	280
M	38,6	44,6	15			46,8	44,4	39
N	42,7	59,6	36			33,6	48,4	220
O	36,0	70,9	35			45,0	39,1	209

Ranking by single indicators (1 – 15)

Institution	Intake	Comp	Publ	Teach	Time	A+B	Credits
A	2	3	3	6	4	4	6
B	7	4	7	1	6	8	5
C	3/	1	4	3	3	10	8/
D	8	5	8	4/	1/	13	14
E	1	2	1	4/	1/	5	10/
F	10	12	6	8	7	2	8/
G	14	9	2	7	5	14	4
H	6	6/	9/			7	10/
I	11	6/	9/			12	15
J	13	13	14	2	8	1	12
K	12	8	12/			11	3
L	5	11	5			9	1
M	9	15	15			3	7
N	3/	14	11			15	2
O	15	10	12/			6	13

Nursing: Indicator scores

Institution	Intake	Comp[17]	Publ[18]	Teach	Time	A+B	Credits	No. of stud.
A-Diak.hj	41,7 x	38,1	12	3,26 x	42,3	65,1	70,2 x	787
B-Stord	37,4 x	22,8	7	3,27 x	44,7	23,2	48,6 x	628
C-Gjøvik	36,4 x	38,3	16	3,41 x	40,7	25,8	47,9 x	816
D-Harstad	34,8	29,1	10	3,01	46,3	27,5	56,6 x	524

---

[15] Faculty of Arts, Social Science and Teacher Education

[16] This figure must be wrong. It may refer to teaching hours?

[17] Figures for relevant department for universities; for university colleges data is only available at faculty level.

[18] Ibid.

E-Hedm	36,5 x	17,7	1	3,45 x	39,5	42,7	51,8 x	792
F-Narvik	38,8 x	51,6	12	3,34 x	45,6	34,7	50,5 x	140
G-Levang	38,0 x	32,6	5	2,98 x	36,9	39,1	46,9 c	822
H-Oslo	39,9 x	41,1	121	2,84 x	36,8	44,7	16,3	3375
I-Sør-Trø	43,0 x	30,1	18	3,32 x	38,8	26,7	53,6 x	1268
J-Telem	39,3 x	28,4	29	3,11 x	41,8	38,8	55,1 x	906
K-Åles	37,2 x	23,1	9	3,13 x	38,9	34,9	50,6 x	585
L-Østf	39,7 x	45,7	26	3,15 x	40,7	18,3	50,0 x	798
M-Sogn	37,2 x	25,8	15	3,89 x	43,4	28,2	53,8 x	685
N-Diakono	38,9 x	28,1	4	3,38 x	39,0	23,5	55,0 x	565
O-Tromsø	39,4 x	28,9	32	3,36 x	38,4	30,2	51,0 x	760
P-Agder	41,0 x	47,1	21	3,20 x	40,6	39,0	47,8 x	1085
Q-Nordl	37,2 x	52,0	41	3,59 x	41,9	35,3	43,1	744
R-Stava	39,8 x	51,1	47	2,78 x	34,6	30,2	47,1 x	874

Ranking by single indicators (1 – 18)

Institution	Intake	Comp	Publ	Teach	Time	A+B	Credits
A	2	8	11/	10	5	1	1
B	12	17	15	9	3	17	12
C	17	7	9	4	8/	15	13
D	18	11	13	15	1	13	2
E	16	18	18	3	11	3	7
F	10	2	11/	7	2	9	10
G	11	9	16	16	16	4	16
H	4	6	1	17	17	2	18
I	1	10	8	8	14	14	6
J	8	13	5	14	7	6	3
K	13//	16	14	13	13	8	9
L	6	5	6	12	8/	18	11
M	13//	15	10	1	4	12	5
N	9	14	17	5	12	16	4
O	7	12	4	6	15	10/	8
P	3	4	7	11	10	5	14
Q	13//	1	3	2	6	7	17
R	5	3	2	18	18	10/	15