

Edith Braun | André Donk | Margret Bülow-Schramm (Hrsg.)

AHELO goes Germany?

Dokumentation des GfHf- & HIS-HF-Workshops

HIS: Forum Hochschule
2 | 2013

HIS  Hochschul
Informations
System GmbH

 **GFHF**
GESELLSCHAFT FÜR
HOCHSCHULFORSCHUNG

Dr. Edith Braun
Tel.: +49 (0)511 1220-477
E-Mail: braun@his.de

Dr. André Donk
Tel.: +49 (0)511 1220-508
E-Mail: donk@his.de

Prof. Dr. Margret Bülow-Schramm
Tel.: +49 (0)40-42838-3836
E-Mail: buelow-schramm@uni-hamburg.de

HIS Hochschul-Informationen-System GmbH
Goseriede 9 | 30159 Hannover | www.his.de
März 2013

Vorwort

Goethe hat in seinen Aphorismen vor fast 200 Jahren das Problem vorweggenommen: „Gewisse Bücher scheinen geschrieben zu sein, nicht damit man daraus lerne, sondern damit man wisse, dass der Verfasser etwas gewusst hat“. Wir möchten provozierend hinzufügen: Schaut man sich die hochschulische Lehre an, so hat man noch oft den Eindruck, dass die Veranstaltungen eher der Präsentation von kanonischem Wissen der Lehrenden dienen als dem Kompetenzerwerb der Studierenden. Sicherlich, in den letzten Jahren ist mit dem Qualitätspakt Lehre und zahlreichen Initiativen im Bereich der Hochschuldidaktik auf diesem Feld vieles erreicht worden. Doch bis heute wissen wir zu wenig verlässlich, mit welchen Kompetenzen und vor allem welchen Kompetenzerwerb die Studierenden die Hochschulen verlassen. Wir möchten deshalb mit diesem Band den wissenschaftlichen Diskurs über die Ergebnisse von Hochschulbildung in Deutschland intensivieren.

Auf dem Workshop „AHELO goes Germany?“, der im November 2012 in Hannover stattfand, wurde ergebnisoffen und mit wissenschaftlichem Anspruch darüber diskutiert, welche Vor-, aber auch welche Nachteile AHELO bei der Erfassung von „learning outcomes“ im Hochschulbereich bietet und was wir für unsere Befassung mit diesem Gegenstand von AHELO lernen können. Der vorliegende Band dokumentiert die Diskussionen und Ergebnisse des hochkarätig besetzten Workshops. Die Worte des großen Hannoveraner Forschers und Intellektuellen bildeten das Motto unserer Anstrengungen: „Die Wege zur Erkenntnis sind interessanter als die Erkenntnis selbst“, so hat Gottfried Wilhelm Leibniz einmal geschrieben. Wir hoffen, dass diese kleine Ausgabe ein wenig zur Aufklärung der „black box“ Kompetenzmessung im Hochschulbereich beitragen kann.

Edith Braun, André Donk & Margret Bülow-Schramm
Hannover & Hamburg, im März 2013

Inhalt

Einleitung

von Margret Bülow-Schramm & Edith Braun..... 1

AHELO im Kontext der Erfahrungen der(hoch)schulischen Kompetenzforschung

AHELO: Norway's Experience of Participation in the Feasibility Study

von Elisabeth Hovdhaugen 7

Ja, aber ... – Lehren aus TEDS-M und KoKoHs für eine Teilnahme an AHELO

von Sigrid Blömeke..... 13

Outcome-Orientierung bei internationalem Vergleich von Studiengängen:

Was kann AHELO von Tuning lernen?

von Margarete Schermutzki & Robert Wagenaar 21

Zur Anlage von AHELO

Vor- und Nachteile Internationaler Vergleiche im Kontext des AHELO-Projekts

von Dominic Orr 27

Chancen und Grenzen von Rankings: Stand der Forschung und Bezug zu AHELO

von Barbara M. Kehm..... 33

AHELO als Ranking? Ergebnisse der Diskussion

von Susanne In der Smitten..... 37

„AHELO goes Germany“ – aber nur, wenn Ziele geklärt, Methoden optimiert und Risiken eingedämmt sind

von Frank Ziegele 41

AHELO im Spannungsfeld von Wissenschaft, Gesellschaft und Bildungspolitik

AHELO: Paradigmenwechsel oder mehr vom Gleichen?

von Stefan Hornbostel 47

AHELO – Zwischen Wissenschaft und Politik

von Hans Pechar..... 55

Kompetenzmessung im Hochschulbereich. Wege aus der Blackbox –

Ein Interview mit Stefan Hornbostel und Edith Braun

von Tanja Meister & André Donk..... 59

Ergebnisprotokoll zum AHELO-Workshop der Gesellschaft für Hochschul- forschung (GfHf)..... 67

Autor(inn)enverzeichnis..... 71

Einleitung

Margret Bülow-Schramm & Edith Braun

In diesem Band möchten wir die Ergebnisse eines Workshops der Gesellschaft für Hochschulforschung, der in Kooperation mit dem HIS-Institut für Hochschulforschung veranstaltet wurde und am 26.11.2012 in Hannover stattgefunden hat, zusammenfassen. Anlass des Workshops war die Ambiguität des OECD-Projekts AHELO (Assessment of Higher Education Learning Outcomes). Auf der einen Seite gibt es Befürworter, die hochschulische Lernergebnisse (learning outcomes) messbar und über Ländergrenzen hinweg vergleichbar machen wollen. Auf der anderen Seite stehen Kritiker, die eine Anpassung der hochschulischen Lehre an die Messinhalte und damit einhergehende negative Folgen erwarten. Ziel war daher insbesondere die wissenschaftlichen Implikationen der AHELO-Studie zu diskutieren.

AHELO – das steht für „Assessment of Higher Education learning outcomes“. Ein Projekt, das die OECD 2009 auf den Weg gebracht hat (Tremblay, Lalancette & Roseveare, 2013). Laut Selbstbeschreibung ist AHELO ein internationales Assessment, das ein robustes Forschungsdesign zur Erfassung des Kompetenzniveaus von Bachelorstudierenden in ihrem letzten Studienjahr entwickeln will. Die entwickelten Instrumente sollen dabei über Sprach- und Kulturgrenzen hinweg valide sein und an den verschiedensten Institutionalisierungsformen hochschulischer Bildung zum Einsatz kommen können. AHELO will ein indikatorengestütztes Monitoringsystem etablieren, das den Hochschulen Daten und Kennwerte übermittelt, die diese dann nutzen können, um ihre Lehre zu verbessern oder gute Lehrergebnisse aus einem Bereich in einen anderen zu transferieren, oder auch die Faktoren guter Lehre auszumachen. Die Teilnahme an AHELO erfolgt dementsprechend freiwillig auf der Ebene der einzelnen Hochschule. Bislang gibt es kein Commitment der deutschen Bildungspolitik, sich an AHELO zu beteiligen.

Spätestens die ersten Befunde der PISA-Studie haben die Aufmerksamkeit in der Diskussion um Bildung auf die Qualität und Resultate von Bildungsprozessen gelenkt. Denn ein zentrales Kriterium für Qualität und Leistungsfähigkeit von Bildungssystemen sind die so genannten „learning outcomes“, also die Ergebnisse von Bildung, die sich u. a. in Lerngewinnen und Lernzuwächsen niederschlagen (Braun, 2011). Für die Gesellschaft und die Politik ist die valide Erfassung der Resultate von Bildungsprozessen so wichtig, dass sie erste dauerhafte Monitoringsysteme etabliert hat, die fortwährend empirische Kennwerte zum Bildungssystem produzieren und diese in das Bildungssystem zurückmelden. Nur so können frühzeitig Fehlentwicklungen im Bildungssystem erkannt und korrigiert werden. Die bei internationalen Assessments besonders erfolgreichen Länder führen ihr gutes Abschneiden im Kern darauf zurück, dass sie diese Steuerungsprinzipien „lernender Organisationen“ auch im Bildungssystem erfolgreich verankert haben und Bildungsinstitutionen und -prozesse regelmäßig evaluieren (Rahmenprogramm des BMBF zur Förderung der empirischen Bildungsforschung). So erreichen sie die Optimierung der Ergebnisse und Wirkungen von Bildung. Während es also mittlerweile seit mehr als zehn Jahren mit PISA ein zuverlässiges und international vergleichendes Instrument gibt, das die Kompetenzen der Schülerinnen und Schüler erhebt, fehlt ein solches bis heute im Bereich der hochschulischen Indikatorensysteme. Kompetenzvermittlung, -erwerb und -gewinn an den Hochschulen stellen eine black box der deutschen Bildungsforschung dar (s. Blömeke in diesem Band).

Dagegen haben sich in der Hochschulforschung in den letzten Jahren verschiedene Indikatoren-systeme etabliert, die z. B. Kennzahlen zu Ausstattung, Kosten und Leistungen der Hochschulen oder zur wirtschaftlichen Situation der Studierenden erfassen. Prominente Beispiele für solche Indikatoren-systeme, die Kennwerte aus dem Hochschulbereich erheben und melden, sind z. B. auf der nationalen Ebene der von der Kultusministerkonferenz sowie dem Bundesministerium für Bildung und Forschung in Auftrag gegebene Bildungsbericht oder das Hochschulranking des CHE. Auf internationaler Ebene sind es unter anderem Bildung auf einen Blick der OECD sowie EUROSTUDENT, die jeweils einen internationalen Vergleich der Leistung von Bildungs- und Hochschulsystemen ermöglichen. Dazu kommen internationale Hochschulrankings wie das der Shanghaier Jiaotong-Universität. „Learning outcomes“ stellen indes bei all diesen Initiativen eine Leerstelle dar. Die erste Frage des Workshops lautete daher: Müssen die bestehenden Indikatoren-systeme nicht um eines ergänzt werden, das auf den Kompetenzerwerb und die Kompetenzzuwächse von Studierenden abhebt und die Lehr- und Lernpraxen an den Universitäten untersucht?

Und wie könnte ein solches Instrument aussehen? Das ist die zweite Frage, der der Workshop nachgegangen ist. So wichtig die Impulse der PISA-Studien für die Kompetenzforschung in der Schule waren, die hochschulbezogene Kompetenzforschung steckt noch in den Anfängen und für ihre Weiterentwicklung können nicht ohne weiteres die Erfahrungen, Ansätze und Ergebnisse aus der Schulforschung übertragen werden. Denn es stehen andere Kompetenzdomänen (als Mathematik, Lesen und Naturwissenschaftliche Kompetenzen) im Mittelpunkt des Interesses. Zudem ist die unterschiedliche Organisation der Bildungsprozesse zu berücksichtigen. So hat hochschulische Bildung auch den Erwerb professioneller Handlungsfähigkeit zum Ziel, die neben einem hohen Niveau an Fachwissen auch motivationale und soziale Aspekte, genauso wie Werthaltungen als überfachliche Kompetenzen einschließt. Wird dieser weite Kompetenzbegriff und der damit einhergehende hohe Stellenwert von überfachlichen oder Schlüsselkompetenzen ernst genommen und als Leitlinie der Gestaltung von hochschulischen Bildungsprozessen zugrunde gelegt, steht die Hochschullehre, aber auch die Erforschung der durch Hochschullehre generierten „learning outcomes“ vor großen Herausforderungen. Erste Erkenntnisse aus der Forschung der Hochschulen belegen, dass die Rückmeldung von studentischen Kompetenzen an die Lehrenden mit einer verstärkten Studierendenorientierung bei den Lehrenden einhergeht (Nowakowski et al., 2012). Das heißt, die Rückmeldung darüber, was Studierende gelernt haben, geht mit der didaktisch erwünschten Haltung der Lehrenden einher, Lehre interaktiver zu gestalten und die konkrete Anwendbarkeit des Gelernten aufzuzeigen. Damit wird ein zentrales Ziel des Bologna-Prozesses in den Hochschulen implementiert: Der Bachelor stellt einen berufsqualifizierenden Abschluss dar (Qualifikationsrahmen für Deutsche Hochschulabschlüsse).

Warum werden Fragen nach dem Kompetenzerwerb und Kompetenzgewinn durch hochschulische Lehre systematisch und vergleichend eigentlich zurzeit eher zurückhaltend erforscht? Ist es nicht vielleicht auch die Sorge von uns Forscher(inne)n, in unserer Funktion als Hochschullehrende gemessen zu werden? Würde ein vergleichsweise niedriges Kompetenzniveau der deutschen Studierenden nicht auf uns alle zurückfallen, die wir im Gegensatz zu den Lehrerinnen und Lehrern an den Schulen meist noch keine systematische Ausbildung zum Unterrichten erhalten haben? Wir müssen uns dieser Herausforderung stellen, die eine dreifache ist: Erstens muss theoretisch gehaltvoll geklärt werden, was Kompetenz bei Studierenden im jeweiligen Fach und zum jeweiligen Zeitpunkt des Studiums bedeutet. Dies ist darüber hinaus eine normative Frage, die uns in den Diskurs mit der Gesellschaft zwingt. Inwiefern sind junge Menschen am Ende ihres Studiums zur „Employability“ ausgebildet worden? Eine solche Debatte rührt an den Fundamenten

Humboldt'scher Bildungsideale. Zweitens müssen wir testen, mittels welcher Methoden und Instrumente die vorab definierten Kompetenzen valide und reliabel gemessen werden können. Im internationalen Vergleich ist dies umso komplexer, da kulturelle und sprachliche Barrieren die Instrumentengestaltung erschweren (s. a. den Beitrag von Orr in diesem Band). Und drittens müssen wir Wege finden, diese Erkenntnisse in die Gestaltung von Curricula und die Aus- und Fortbildung von in der Hochschullehre tätigen Personen einfließen zu lassen.

Hinsichtlich der Zielstellung der Hochschullehre vollzog sich in der Hochschuldidaktik bereits in den letzten Jahren ein „Shift from Teaching to Learning“, der sich allerdings in der praktischen Lehre selbst nur ansatzweise durchgesetzt hat. Die Kompetenzforschung nimmt sich zunehmend der Frage nach „learning outcomes“ im tertiären Bereich an. Und es finden mittlerweile an jeder Hochschule punktuelle, z. T. systematische Evaluationen zur Lehrqualität statt. Aber weithin akzeptierte Tests mit standardisierten Instrumenten, die nationale wie internationale Vergleich erlauben – und damit Kennwerte erst wirklich aussagefähig machen – fehlen bis dato. Aufgrund dieser Forschungsdesiderate gilt es zu klären, ob das OECD-Projekt AHELO die Antwort sein kann.

Viele Kompetenzforscherinnen und -forscher sind sich zum jetzigen Zeitpunkt einig: Das geeignete Instrument zur Testung im Hochschulstudium erworbener Kompetenzen ist noch nicht entwickelt, viele konzeptuelle und definitorische Vorarbeiten sind noch notwendig, bis die Testung selbst vorgenommen werden kann. Es gilt zudem nicht nur aktuelle Kompetenzniveaus, sondern auch Lernzuwächse auszumachen. Das heißt, es muss auf der Individualebene – unter Berücksichtigung der Eingangsbedingungen – ein Längsschnittvergleich von Kompetenzen erfolgen. Da in der Hochschule bislang derartige Instrumente fehlen, kann von AHELO hier ebenfalls ein Impuls ausgehen.

Die Erwartungen, die in AHELO gesetzt werden, sind entsprechend hoch und auch jene Institutionen, die sich aktuell kritisch gegenüber Rankings gezeigt haben, signalisieren Interesse am Projekt. Dieter Lenzen, der Präsident der Universität Hamburg, sagt in einem Interview mit den Potsdamer Neuesten Nachrichten vom 2. Oktober 2012: „Bei einem Forschungsprojekt der DFG über Universitäten würden wir dagegen immer davon ausgehen, dass es methodisch okay ist, und mitmachen. Ein anderes Beispiel: Wenn das AHELO-Projekt der OECD tatsächlich durchgeführt wird, könnten interessante Erkenntnisse zu erwarten sein. Kernfragen sind doch: Was muss eine Uni im Unterricht leisten? Wie und wo können Studierende ihre Wissensstrukturen am besten verändern? AHELO könnte vielleicht Antworten darauf geben, ob und wie wir das messen können“.

Siebzehn Länder, darunter unter anderem Australien, Belgien, Japan, Korea, Mexiko, die Niederlande oder Norwegen, haben an der AHELO-Machbarkeitsstudie, die im letzten Jahr durchgeführt wurde, teilgenommen. In dieser Machbarkeitsstudie sollte untersucht werden, ob ein solch internationales Assessment wissenschaftlich ertragreich und forschungsorganisatorisch praktikabel umgesetzt werden kann. Auch ist nach wie vor ungeklärt, wie ein nationaler Vergleich geleistet werden kann, wenn die Untersuchung auf der Ebene einzelner, freiwillig teilnehmender Institutionen erfolgt. Handelt es sich nicht vielmehr um einen Vergleich von Hochschulen als von Hochschulsystemen? Denn die Erfahrungen haben gezeigt, dass die Konstruktion und Anwendung eines solchen Assessments kein triviales Unterfangen ist. Die Ergebnisse der Machbarkeitsstudie, darauf deuten öffentliche Äußerungen hin, werden wahrscheinlich nicht eindeutig für eine schnelle, breite Durchführung von AHELO sprechen. So lässt sich Professor Richard James vom Centre for the Study of Higher Education der Universität Melbourne jüngst in der Zeitschrift *The Australian* mit den Worten zitieren: „The spotlight is squarely on the OECD and whether they can deliver on this ambitious project, and it would be fair to say the scepticism levels are high“ (Lane,

2012). Daher ist zum heutigen Tage noch offen, ob und in welcher Form AHELO weitergeführt wird und ob und in welchem Maße sich Deutschland beteiligen kann und wird.

Der Band möchte vor diesem Hintergrund und mit einem dezidiert wissenschaftlichen Anspruch einen Überblick über die derzeitigen Potenziale der AHELO-Studie geben sowie über aktuelle Entwicklungen und Erkenntnisse der hochschulischen Kompetenzforschung informieren. Die Struktur des Bandes weicht von der Chronologie der Workshops ab und ordnet die Beiträge unter drei inhaltlichen Gesichtspunkten.

Im ersten Teil des Bandes „AHELO im Kontext der Erfahrungen der (hoch)schulischen Kompetenzforschung“ stellt Elisbeth Hovdhaugen erste Erfahrungen und Befunde aus der norwegischen Praxis der Machtbarkeitsstudie vor. Sigrid Blömeke bespricht anschließend Parallelen zu der TEDS-M (Teacher Education and Development Study in Mathematics) Studie und arbeitet die Notwendigkeit eines methodisch fundierten wie reflektierten Vorgehens bei der Erfassung hochschulischer Kompetenzen und Kompetenzentwicklung heraus. Margret Schmermutzki und Robert Waagenar stellen das TUNING-Projekt vor, das den konzeptionellen Rahmen für AHELO gebildet hat.

Die Beiträge im zweiten Teil „Zur Anlage von AHELO“ hinterfragen kritisch die Konzeptualisierung des AHELO-Projekts. Zunächst bearbeitet Dominic Orr die Möglichkeiten und Grenzen des internationalen Vergleichs als Methode der Hochschulforschung. Barbara M. Kehm geht in ihrem Beitrag der Frage nach der Bedeutung von Rankings im Hochschulbereich mit Bezug auf AHELO und den mit diesen verbundenen Chancen und Problemen nach. Anschließend dokumentiert Susanne in der Smitten die Diskussionen des Workshops hinsichtlich der aufgeworfenen Rankingproblematik. Frank Ziegele erörtert Möglichkeiten von AHELO als „Lerninstrument“ für Hochschulen im Kontext bestehender Rankings.

Der dritte Teil des Bandes widmet sich den Implikationen von „AHELO im Spannungsfeld von Wissenschaft, Gesellschaft und Bildungspolitik“. Stefan Hornbostel fragt, inwiefern AHELO einen Paradigmenwechsel in der Gestaltung tertiärer Ausbildung einläuten kann, und Hans Pechar wird in seinem Vortrag das Spannungsfeld von Wissenschaft und Politik aufzeigen, in dem sich Assessments im Hochschulbereich bewegen. Abschließend diskutieren Stefan Hornbostel und Edith Braun Wege aus der Blackbox der Kompetenzforschung.

Literatur

Braun, E. (2011). Hochschulforschung in der Psychologie. *Die Hochschule. Journal für Wissenschaft und Bildung*, 2/2011, 74–90.

Bundesministerium für Bildung und Forschung (2007). Rahmenprogramm zur Förderung der empirischen Bildungsforschung. Online-Dokument: http://www.empirische-bildungsforschung-bmbf.de/_media/Rahmenprogramm.pdf

Lane, B. (2012). Doubts over OECD comparison. In: *The Australian*, 10.10.2012. Online-Dokument: <http://www.theaustralian.com.au/higher-education/doubts-over-oecd-comparisons/story-e6frgcjx-1226492348956>

Lenzen, D. (2012). Rankings sind Unfug. In: *Potsdamer Neueste Nachrichten*, 02.10.2012, S. 24.

Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2005). Qualifikationsrahmen für Deutsche Hochschulabschlüsse. Online-Dokument: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2005/2005_04_21-Qualifikationsrahmen-HS-Abschluesse.pdf

Tremblay, K.; Lalancette, D.; Roseveare, D. (2013). AHELO. Feasibility Study Report. Paris: OECD.

AHELO im Kontext der Erfahrungen
der(hoch)schulischen Kompetenzforschung

AHELO: Norway's Experience of Participation in the Feasibility Study

Elisabeth Hovdhaugen

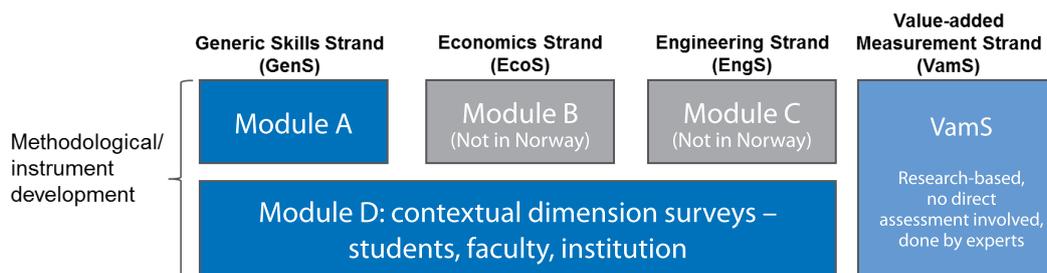
The aim of this paper is to give an overview of Norway's experience in participating in the AHELO feasibility study, which was conducted in 2010-2012 and will be reported in 2013. At the time of writing the reporting stage has not yet been concluded and the paper therefore describes the process, and Norway's experience of being part of the international feasibility study.

Introduction to AHELO as a project

AHELO stands for Assessment of Higher Education "learning outcomes", which is an initiative from the Organisation for Economic Cooperation and Development (OECD) to assess what students know and can do upon graduation. It was introduced by a feasibility study which ran from 2009 until 2012, with the aim of testing concepts and the potential implementation of a test of "learning outcomes", to be administered across institutions in different countries.

The project was organised in several strands: a Generic skills strand aiming at testing critical and analytical thinking, and two disciplinary strands, in Economics and Engineering. In addition the project included contextual surveys, aiming at collecting information on students, faculty and the institution to facilitate comparison across institutions. The final strand was the value-added measurement strand, which was not included as a part of the feasibility project, but rather as a desk-based overview of research done by experts. An illustration of the different strands in the project is presented in figure 1.

Figure 1: Over-view of strands in the AHELO project



There are a large and diverse group of countries involved in AHELO, with Norway, Finland, Korea, Mexico, Egypt and some states in the US, among others. This made it clear right from the start that there was rich potential for comparing contrasting higher education institutions across cultures and disciplines. However, it is important to clarify that AHELO focuses on institutions and not on countries. Hence, AHELO aims at providing institutions with information on students' "learning outcomes", and an option to compare themselves to similar institutions in other countries; the test should not just produce another international ranking, but should instead be a way for institutions to mirror themselves against similar institutions.

In light of this, the test had to take into account several layers of difference; the differences between national systems of higher education and the diversity of higher education institutions, both within a country and between countries, in addition to variations in of the degree of selection of students in the system or in certain higher education institutions, variations in the duration and content of programmes as well as cultural and linguistic diversity. All of this was to be covered by the context surveys (module D).

The aims of AHELO are two-fold: firstly to test whether a reliable cross-national comparison of higher education outcomes is scientifically possible, and secondly to test whether a valid assessment can be practically implemented within institutions, in the countries that participated.

Norway's participation in AHELO

Norway became involved in the AHELO Feasibility study at the start of the project in 2009. The Ministry of Education decided that Norway would participate in one of the three strands of the project: the Generic skills strand. The aim of the Generic skills strand was to measure types of learning outcome or skills that would be expected across all kinds of higher education programmes, such as various kinds of analytical thinking, coherently communicating a conclusion or argument, as well as some common knowledge e.g. interpreting specific kinds of data presented in graphs/charts.

Types of tests used in AHELO

The Generic skills test consisted of two types of test, a Constructed Response Task (CRT) and a set of Multiple Choice Questions (MCQs). In the CRT the test-taker is presented with a hypothetical situation where they are supposed to give a recommendation or present a viable explanation for a phenomenon, using a given set of digital documents. Their answer has to be presented as a form of essay. The MCQs are structured with a text-based or figure-based introduction, followed by 2-5 multiple choice questions (items), all with four alternative answers. In total, the MCQ-test contains 25 items. Some types of MCQs also contain mathematical formulas that the test-taker has to use. However, the questions demand the test-taker to analyse a problem or interpret the text/figure/formula in order to find which alternative is right. Both tests share the aim of testing analytical thinking or the application of information provided to solve a problem, rather than testing general knowledge about the content/topics dealt with in the questions. In addition, the CRT aim at testing student writing skills.

In addition to the tests, there were three context instruments, which were surveys collecting relevant context and background information about the students taking the test, faculty at the institution and the institution. The idea behind collecting context information at the student, faculty and institutional level was to provide a means to control for differences between institutions, and a resource to help explain institutional differences in test results.

Based on common principals of testing, and in line with similar international tests such as PISA, both types of tests were randomized, meaning that not all questions were answered by all participants. There were two CRTs, and students were randomly assigned to one of them. The MCQ had a core-part (15 questions) that all test-takers had to complete, in addition to four sets of 10 questions which were rotated (so a quarter of students took each set of additional questions).

Norway's experience in AHELO

The AHELO project was divided into two phases; the first phase consisted of preparation such as making translations and adaptations of the two tests in addition to the context instruments, while the second phase consisted of the implementation of the study. The tests used in the Generic skills strand in the AHELO feasibility study were based on an American CRT and MCQs originally developed in Australia. Hence, in both cases the original language of the tests was English. The implementation phase consisted of several procedures, such as sampling, testing the online AHELO system/tools at institutions, recruiting students to participate in the assessment and running the test at institutions, as well as getting faculty and institution coordinators to complete the context survey. The student context survey was distributed as a part of the main AHELO assessment.

Translations

The translations of the CRT started at the beginning of the first phase while the translation of MCQs and context instruments started after the decision had been taken to go on to the implementation phase, in July 2011. Due to this, there was plenty of time for translation of the CRT while the other translations were time-constrained and had to be done in haste. Despite this, the translations of the MCQs came out quite well, due to good routines for feedback and review between the consortium and national team. Translations of the CRT were also good, although these proved to be time-consuming and costly, as the CRT involves a digital library with 8-10 documents for each of the two versions of the test, which generated a lot of documents to be translated. There were also specific challenges related to translating the scoring instrument that had to be used to score the CRT. The scoring rubric contained many nuances in English which could not directly be translated into Norwegian. In comparison, the MCQs proved more straight-forward and cost-effective to translate, and these tests are also scored automatically, which contributed to making them even more cost effective.

Sampling

Since this was a feasibility study, with no intent of measuring and comparing at the higher education system level, a convenience sample of institutions was recruited in all countries. In Norway five institutions participated in the AHELO Feasibility study. The participating institutions were quite different from each other, both in size and mission. It was not hard to get institutions to volunteer to participate in AHELO, but it was challenging to keep institutions engaged, since the project period was longer than originally anticipated.

The aim of AHELO was to sample students in their third or fourth year, at the end of their bachelor's degree. In Norway that definition was extended to include students that had started an integrated master's programme and were at the end of the bachelor's part of their degree (in their third year). Students were sampled randomly across each institution, to give a sample of 300 students per institution, regardless of the size of the institution. The sample was drawn from the student records at the participating institutions, using the same definitions in all cases. The students were sampled across disciplines and programmes, mirroring the disciplinary composition of the institution. The AHELO consortium had provided a quite extensive sampling manual, that specified who should be included and excluded in the sample.

Faculty were also randomly sampled across disciplines, and the data about the faculty population was also collected through institutional records. However, the design was not intended to link students and faculty, so there was no connection between the student and the faculty sample. The faculty survey was short, only 15 questions distributed to faculty via email.

Implementation

Norway was the first country to start testing, in early February 2012. However, due to late delivery of log-in codes, one week of testing had to be cancelled at the institution that was to start first, which may have led to a minor loss of participants. However, early on in the testing process it also became evident that it was challenging to recruit enough students to take the test. A key reason for this is thought to be the duration, with the Generic skills test taking over two hours to complete.

Students were recruited using many different approaches. All sampled students received an email invitation to take part and could sign up for a testing session of their choice. They then had to meet up at the computer lab to take the test on the day they had chosen. Institutions also used a range of measures to inform students about the opportunity to participate in AHELO, but most sampled students chose to not take part. Institutional coordinators tried a lot of different measures, such as information online and flyers around campus, articles in campus newspapers and direct information to students in classes, and worked hard to provide information and encouragement in the weeks before testing. All institutions also tried different forms of incentives, but they all seemed to have little or no effect.

There were few technical problems in implementing the test, but it also proved essential that the institutions had the opportunity to try out the testing systems beforehand, in order to find and resolve potential problems. For example, it proved quite time-consuming to set up the computers to provide a secure test environment (blocking access to other websites during the test to avoid cheating) and this procedure had to be put in place and tested in order to function well during the actual test. However, during the test sessions there were few technical problems, and as such the testing went smoothly.

International collaboration

The international collaboration was good with regards to research tools and processes, extensive efforts were also made so that all countries would use the same methods and maintain comparability across institutions. However, there were challenges related to diverging semester dates and structure of study programmes in some countries. These challenges were relevant to recruitment of students in Norway, as a lot of students in their last year of a bachelor's degree are concentrating on writing their bachelor's thesis and some are even not on campus. There may also be differing relationships between higher education institutions and their students in different countries, and this could explain some of the variation in response rates; in some countries being chosen to participate in such an international study might be seen as an honour or an obligation towards one's institution, but in Norway students may be less likely to consider participation in such terms, and did not seem to prioritise participation. Some students also commented that they saw little purpose in taking a test that did not provide them with a grade or a credit.

Contributions so far

AHELO has received considerable attention from governments and institutions interested in the area of “learning outcomes”, and this has increased over time. However, it seems as if one of the most important groups of participants, the students, were overlooked to some extent in the preparation and approach taken in AHELO; it was assumed that they would volunteer their time to participate in the test, but few did. At the same time, given that AHELO was not designed to give students feedback on their own performance, or to award credits towards a degree, there seemed to be little clear rationale for student participation. Sadly, AHELO was not as attractive to students as the OECD had hoped it would be, at least in some countries.

However, one of the main contributions of AHELO is the attempt to measure “learning outcomes” directly, which has not been done in a cross-country context before. Working with AHELO has shown the importance of a common framework, an agreement on what the tests are supposed to measure. This was not in place in the Generic skills strand, and has been a hindrance in the process. The first report on AHELO from the OECD also indicates that any future AHELO has to start up with agreeing on a common framework for the skill measured.

There are also some practical challenges that remain for AHELO, and the biggest challenge is probably how to get high response rates and how to engage students in the study. In order to do so, the study has to offer something to students to make them want to volunteer their time.

Another key aim of AHELO assumes that it is possible to make ‘valid’ comparisons at the institutional level. In the feasibility study it might be hard for institutions to find a “benchmark” institution or group of institutions, and questions remain about this intention, even extensive context instruments can provide sufficient controls for this. This remains to be seen once the data has been fully analysed. In order for AHELO to make an important contribution to the discussion on “learning outcomes” we have to make sure that the data can provide more information to institutions than grades and other simpler outcome measures already do.

Ja, aber ... – Lehren aus TEDS-M und KoKoHs für eine Teilnahme an AHELO

Sigrid Blömeke

Der tertiäre Bildungsbereich ist in weiten Teilen noch immer eine *Blackbox*, was die objektive und standardisierte Erfassung der Leistungen seiner Absolventinnen und Absolventen angeht. Während unser Schulsystem als Ganzes und die einzelnen Schulen seit rund 10 Jahren national und international-vergleichend in vielfacher Form auf den Prüfstand gestellt werden – sei es durch Large-Scale-Assessments wie TIMSS, PIRLS oder PISA, sei es durch Vergleichsarbeiten oder zentrale Abschlussprüfungen, sei es durch regionale Schulinspektionen –, erfreuen sich Universitäten, Pädagogische Hochschulen, Fachhochschulen und die weiteren tertiären Bildungseinrichtungen einer bemerkenswerten Freiheit. Auch an der Pilotstudie zum „Assessment of Higher Education „learning outcomes“ (AHELO)“ der OECD hat sich Deutschland nicht beteiligt.

Ausgerechnet den teuersten Bildungsbereich, in dem die gesellschaftliche Elite geformt werden soll, entsprechenden Qualitätsprüfungen zu entziehen, erscheint wenig zielführend. Im Kontext einer globalisierten Wissensgesellschaft, in der die Anforderungen an Akademikerinnen und Akademiker einem ständigen Wandel unterliegen und immer komplexer werden, müsste dem tertiären Bildungssektor in der empirischen Evaluationsforschung eher mehr als weniger Aufmerksamkeit geschenkt werden. Nur auf der Basis von Wirksamkeitsprüfungen können evidenzbasiert Aussagen zur Leistungsfähigkeit der Hochschullandschaft gewonnen und für nachhaltige Entwicklungs- und Optimierungsmaßnahmen auf struktureller, organisatorischer und individueller Ebene genutzt werden. Die Erfahrungen aus den Umstrukturierungsprozessen der Vergangenheit (z. B. der Umstellung auf das Bachelor-Master-Modell) zeigen, dass diese einer formativen und summativen Evaluation bedürfen.

Unabhängig von der Gewinnung bildungspolitisch relevanten Steuerungswissens durch solche Studien – die Bedeutsamkeit eines solchen Ziels kann normativ unterschiedlich beantwortet werden – würde sich die Hochschulforschung ohne ein stärkeres Engagement in der Kompetenzforschung in jedem Falle einer reizvollen Forschungsperspektive berauben. Im Detail wissen wir über die Eigenschaften akademisch erworbener Kompetenzen so gut wie nichts: Wie sind diese inhaltlich strukturiert, welche Leistungsniveaus lassen sich identifizieren, entwickeln sich akademische Kompetenzen in Stufen? Welche Rolle spielen kognitive Facetten, welche Bedeutung kommt affektiv-motivationalen Facetten zu und welche Eigenschaften differenzieren besonders gut zwischen erfolgreichen und weniger erfolgreichen Absolventinnen und Absolventen? Die bisherige Hochschulforschung beruht fast vollständig auf Befragungen, mit denen sich solche Erkenntnisse nicht gewinnen lassen (Braun, 2011).

Allein die Notwendigkeit, sich über einen theoretischen Rahmen, Studiendesigns und Instrumente für Kompetenzmessungen bei Studierenden zu verständigen, würde zu zahlreichen interessanten Forschungsfragen führen, die in inhaltlich, institutionell und methodisch vielfältige Projekte münden könnten (Beispiele hierfür sind ebenso wie der entsprechende Forschungsstand zusammengefasst in Blömeke et al., 2013; Zlatkin-Troitschanskaia & Kuhn, 2010). Die Mehrdimensionalität akademisch erworbener Kompetenzen von Studierenden und Promovierenden sowie die inter- und intranationale Vielfalt an Studienmodellen, Ausbildungsstrukturen und Lehrleistungen stellen hohe Ansprüche an die Forschungsmethodik. Triangulationen unterschiedli-

cher Zugänge und Weiterentwicklungen des methodischen Status Quo wären unumgänglich, sodass auch in dieser Hinsicht Fortschritte stattfinden würden.

Insofern stellt der vorliegende Beitrag ein Plädoyer für mehr Kompetenzforschung im Hochschulsektor und zugleich für eine Teilnahme Deutschlands an AHELO dar. AHELO und weitere Studien könnten auf den in Deutschland bereits vorhandenen Ansätzen aufbauen und die dort erkannten Probleme adressieren. Um zum Erfolg zu werden, stellt eine präzise Klärung der Ziele, Rahmenbedingungen und Grenzen solcher Studien allerdings eine Voraussetzung dar. Als federführend für eine entsprechende Forschung können in Deutschland die internationale Vergleichsstudie TEDS-M und die Förderinitiative KoKoHs angesehen werden. Die beiden Vorhaben lassen aber auch die Probleme der Kompetenzforschung erkennen und was noch geleistet werden muss, bevor mit Vorhaben wie AHELO ein bedeutsamer Erkenntnisgewinn verbunden wäre. Im Folgenden werden die beiden Studien zunächst kurz vorgestellt, bevor zentrale Fragen herausgearbeitet werden, die vor dem Beginn weiterer Studien jeweils geklärt werden müssen.

Vorhandene Ansätze der Kompetenzmessung im Hochschulsektor I: TEDS-M

Die internationale Vergleichsstudie „Teacher Education and Development Study: Learning to Teach Mathematics (TEDS-M)“ wurde von der „International Association for the Evaluation of Educational Achievement (IEA)“ durchgeführt, die eine Nicht-Regierungsorganisation ist. Die IEA ist ein freier Zusammenschluss von Wissenschaftseinrichtungen. Für Deutschland agieren in ihrem Rahmen z. B. das DIPF in Frankfurt/Main, die Humboldt-Universität zu Berlin und das Institut für Schulentwicklungsforschung in Dortmund. Die IEA führt Studien wie TIMSS, PIRLS oder die CIVIC Education Study durch.

Aufbauend auf der Pilotstudie „Mathematics Teaching in the 21st Century (MT21)“ (Schmidt, Blömeke & Tatto, 2011) war TEDS-M das erste Large-Scale-Assessment im Hochschulsektor, das mit national repräsentativen Stichproben und direkten Kompetenztests arbeitete (Tatto et al., 2008, 2012). Fast 24.000 angehende Primar- und Sekundarstufe-I-Lehrkräfte von 800 Institutionen aus 16 Ländern, die eine Lehrberechtigung für den Mathematikunterricht in den Klassen 1 bis 4 bzw. 8 erwerben würden, wurden im letzten Jahr ihrer Ausbildung auf ihr mathematisches, mathematikdidaktisches und allgemeinpädagogisches Professionswissen getestet sowie zu ihren Überzeugungen und Lerngelegenheiten befragt. Bei den Ländern handelte sich um Botswana, Chile, Deutschland, Georgien, Kanada, Malaysia, Norwegen, den Oman, die Philippinen, Polen, Russland, die Schweiz, Singapur, Spanien, Taiwan, Thailand und die USA.

Da mit Ausnahme von Thailand und Malaysia in allen übrigen Teilnahmeländern für die Primarstufe Klassenlehrkräfte ausgebildet wurden, war jeweils die gesamte Grundschullehrerausbildung in TEDS-M einbezogen. Mathematik zu unterrichten stellt für diese Lehrkräfte eine kleine, aber gesellschaftlich bedeutsame Aufgabe dar. In der Sekundarstufe I dominiert das Fachlehrerprinzip, sodass angehende Lehrkräfte mit dem Unterrichtsfach Mathematik untersucht wurden (detaillierte Informationen zur Studie und ihren Ergebnissen siehe z. B. in Blömeke et al., 2012a, b)

Von den angehenden Grundschullehrkräften zeigten jene aus Taiwan am Ende der Ausbildung das höchste mathematische Professionswissen. Starke Leistungen zeigten auch die Lehrkräfte aus Singapur, der Schweiz, Russland, Thailand und Norwegen. Auch im Bereich der Mathematikdidaktik wurde die Leistungsspitze von den Lehrkräften aus Singapur und Taiwan gebildet. Über dem internationalen Mittelwert lagen erneut die mathematikdidaktischen Leistungen angehender Grundschullehrkräfte aus Norwegen und der Schweiz sowie aus den USA.

TEDS-M ordnete Deutschland im Mittelfeld der Teilnahmeländer ein und gab wichtige Hinweise auf mögliche Ursachen der Schwächen Deutschlands im Bereich des schulischen Lernens. Vertiefende Untersuchungen zeigen, dass für das erreichte Ergebnis am Ende der Grundschullehrerausbildung entscheidend ist, in welchem Umfang den angehenden Lehrkräften fachbezogene Lerngelegenheiten geboten werden. In den Ländern an der Spitze der TEDS-M-Verteilung liegt der Anteil an Mathematik in Breite und Tiefe deutlich höher als in den Ländern, die schwächer abschneiden.

Selbst wenn die demographischen und schulischen Voraussetzungen der in die Lehrerausbildung eintretenden Personen kontrolliert werden, sodass Leistungsunterschiede nicht auf Rekrutierungsvorteile beispielsweise in Ländern wie Taiwan zurückgeführt werden können, wo der Lehrerberuf hoch angesehen ist, ist der Mehrwert umfangreicher Fachangebote deutlich erkennbar. Dieses Ergebnis kann innerhalb der an TEDS-M teilnehmenden Länder repliziert werden: Unter Kontrolle von Eingangsvoraussetzungen wie der durchschnittlichen Schulleistung oder dem Umfang an mathematischer Schulbildung weisen Grundschullehrkräfte aus Ausbildungsgängen mit mehr Mathematik und Mathematikdidaktik am Ende ein höheres fachbezogenes Professionswissen auf.

Im Bereich der Sekundarstufe I liegen die Kompetenzen angehender Mathematiklehrkräfte aus Taiwan ebenfalls an der Spitze. Sie verfügen im internationalen Vergleich über das höchste mathematische und mathematikdidaktische Professionswissen. Eine Gruppe von fünf Ländern zeichnet sich dadurch aus, dass ihre angehenden Lehrkräfte in Mathematik und Mathematikdidaktik über dem internationalen Mittelwert liegen. Zu dieser Gruppe gehört Deutschland. Daneben handelt es sich um Russland, Polen, Singapur und die Schweiz.

Für die Sekundarstufe I gilt in noch stärkerem Maße als für die Grundschule, dass unter sonst gleichen Umständen im Hinblick auf den demographischen und schulischen Hintergrund der angehenden Lehrkräfte bedeutsam ist, wie umfangreich die mathematische Ausbildung ist. Diese findet in allen Ländern, die in TEDS-M besonders gute Leistungen erreichten, an Universitäten statt und sie dauert mindestens 4,5 Jahre. Die Konzentration auf nur ein Unterrichtsfach scheint dabei mit Vorteilen verbunden zu sein.

Berücksichtigt man, dass sich die Wirkungskette fortsetzen lässt, indem Lehrkräfte mit höherem mathematischen und mathematikdidaktischem Professionswissen stärker kognitiv aktivierende Unterrichtsmaterialien einsetzen und auf diese Weise höhere Schülerleistungen erzielen (Baumert et al., 2010), wird deutlich, wie bedeutsam eine sorgfältige fachliche Ausbildung für die Leistungsfähigkeit eines Schulsystems ist.

Vorhandene Ansätze der Kompetenzmessung im Hochschulsektor II: KoKoHs

Das Bundesministerium für Bildung und Forschung (BMBF) hat angesichts des Forschungsdefizits zur Kompetenzmessung im Hochschulsektor unter der Leitung von Sigrid Blömeke (Humboldt-Universität zu Berlin) und Olga Zlatkin-Troitschanskaia (Johannes-Gutenberg-Universität Mainz) die Förderinitiative „Kompetenzmodellierung und -erfassung im Hochschulsektor (KoKoHs)“ gestartet. Mehr als 70 Einzelprojekte sind in 23 Projektverbänden und fünf thematisch ähnlichen Clustern gebündelt, die den Kompetenzerwerb in Kernbereichen des Hochschulsektors untersuchen: in den Ingenieur-, Wirtschafts- und Sozialwissenschaften, in der Lehrerausbildung in MINT- und SoWi-Fächern sowie für pädagogische Berufe und generische Kompetenzen. Ziel ist die Entwicklung reliabler und valider Instrumente zur Erfassung kognitiver Leistungsdispositionen, die eine Voraussetzung für die Bewältigung typischer Anforderungen in akademischen Berufen darstellen (Details siehe z. B. in Zlatkin-Troitschanskai et al., 2012).

Ein zentrales konzeptuelles Merkmal von KoKoHs besteht darin, dass ein Fokus auf die Modellierung von domänenspezifischen Kompetenzen gelegt wird. Eine valide Kompetenzerfassung ist nur dann zur Generierung von bildungspolitischem Steuerungswissen geeignet, wenn die spezifischen Rahmenbedingungen mit untersucht werden. Daher ist eine Prozess- und Strukturanalyse des Hochschulbereichs mit seinen institutionellen Varianten, seiner Vielfalt an Fachwissenschaften und Studienstrukturen, seinen divergierenden gesellschaftlichen und wirtschaftlichen Rahmenbedingungen und die in der akademischen Ausbildung gebotenen Lerngelegenheiten (Inhalte und Methoden etc.) zu vollziehen.

Neben Aussagen dazu, welche Anforderungssituationen bei welcher individuellen Kompetenzausprägung zu bewältigen sind, sind neben übergeordneten institutionellen Anforderungen besonders die spezifischen Charakteristika eines Fachs bzw. der Domäne (z. B. curriculare Aspekte, Inhaltsbereiche, situative Einbettung) sowie die mit ihnen verbundenen kognitiven Anforderungsniveaus zu berücksichtigen. Ein erster Ansatz dazu kann im „Qualifikationsrahmen für Deutsche Hochschulabschlüsse“ gesehen werden, in dem zu erwerbende Kompetenzen („Wissen und Verstehen“ sowie „Können“) und formale Aspekte (z. B. Zugangsvoraussetzungen, Dauer) gestuft nach Abschlüssen (Bachelor, Master, Promotion) formuliert werden. Dem Rahmen mangelt es allerdings an einer empirischen Fundierung und am Domänenbezug.

Neben der (Weiter-)Entwicklung von Kompetenzmodellen liegt ein zweiter Förderschwerpunkt von KoKoHs auf methodischen Fragen der Kompetenzmessung. Die Komplexität und Mehrdimensionalität des Forschungsbereichs bedarf einer ebenso komplexen und mehrdimensionalen messmethodischen Umsetzung. Theoriegeleitet entwickelte Aussagen über die Dimensionierung, Graduierung und Entwicklung domänenspezifischer Kompetenzen stellen die notwendige Voraussetzung für die Generierung geeigneter Messinstrumente dar. Um eine adäquate Umsetzung der theoretischen Modelle in Messverfahren zu gewährleisten, bedarf es aber auch der Verwendung adäquater psychometrischer Zugänge. So gilt es neben personenbezogenen (latenten) Merkmalen auch den situativen Anforderungsbezug zu erfassen. Psychometrische Modelle, die diesen Anforderungen gerecht werden, basieren vorzugsweise auf der Item-Response-Theorie (IRT).

Diese kurzen Projektdarstellungen zu TEDS-M und KoKoHs vernachlässigen inhaltliche und methodische Herausforderungen im Detail, mit denen die Projekte konfrontiert waren. Diese werden im Folgenden gesondert diskutiert.

Evaluation von Bildungssystemen vs. Evaluation einzelner Institutionen

Bei internationalen Vergleichsstudien stellt sich immer die Frage, inwieweit die angestellten Vergleiche sinnvoll sind. Werden „Äpfel mit Birnen“ verglichen oder stecken hinter den untersuchten Konstrukten weitgehend einheitliche Rahmenbedingungen, sodass sich valide und weiterführende Schlussfolgerungen ziehen lassen? Die nationale und internationale Variabilität ist im sekundären Bildungsbereich, der in TIMSS und PISA auf dem Prüfstand steht, bereits hoch. Wir haben es innerhalb Deutschlands zum Beispiel mit unterschiedlichen Schulformen sowie für diese wiederum mit unterschiedlichen Curricula je nach Bundesland zu tun. Im internationalen Vergleich kommen Unterschiede in der Struktur der Schulsysteme und unterschiedliche Eingangsvoraussetzungen der SchülerInnen hinzu.

Im tertiären Bildungsbereich potenziert sich diese Variabilität nicht zuletzt aufgrund der fast überall vorhandenen Hochschulautonomie. Hier muss daher genau geklärt werden, welche Art von Studie man will. TIMSS, PIRLS, PISA und letztlich auch TEDS-M waren Studien, die eine Evaluation von Bildungssystemen unabhängig von ihrer konkreten strukturellen Ausgestaltung vor-

nahmen. Aus den Studien ließen sich keine Aussagen zur Qualität und Effektivität einzelner Institutionen ableiten. Das war nicht gewollt und nicht Ziel von TEDS-M, TIMSS etc.

Nur in einer solchen Konzeption, die die Effizienz von Gesamtsystemen beschreibt, indem das am Ende des Bildungsabschnitts erreichte Leistungsvermögen der Gesamtpopulation erfasst wird, ist es sinnvoll, Ländervergleiche anzustellen. Die zwischen den Ländern herrschende Variabilität wird dann als eine bildungspolitische Entscheidung angesehen, die mehr oder weniger effizient ausfallen kann. Die deutsche Bildungspolitik hätte beispielsweise die Möglichkeit gehabt, anstatt eines gegliederten Schulsystems ein Gesamtschulsystem einzuführen oder statt eines föderalen Systems mit vielfältigen Curricula ein zentrales System mit einem einheitlichen Curriculum.

In Bezug auf AHELO würde ein solches systembezogenes Studiendesgin bedeuten, dass für ein bestimmtes Feld – z. B. Ingenieurwissenschaften – zufällig Hochschulen entsprechend der Größe ihrer Studierendenzahl in diesem Feld gezogen würden und innerhalb dieser Hochschulen zufällig gezogene Studierende im letzten Jahr ihrer Ausbildung getestet würden. Deren Ergebnisse würden national aggregiert und mit den aggregierten Ergebnissen anderer Länder verglichen. Dass wir in Deutschland zwischen Universitäten und Fachhochschulen differenzieren, dass die Studiengänge unterschiedlich lang und curricular unterschiedlich gestaltet sind, wären Entscheidungen, die bewusst getroffen wurden und deren Auswirkungen nun auf nationaler Ebene erfasst werden.

Was eine solche Anlage von AHELO bei einer entsprechenden Anlage der Stichprobenziehung neben Ländervergleichen noch ermöglichen würde, wäre der internationale Vergleich von kleineren Gruppen an Studiengängen (z. B. statt die Ingenieurwissenschaften als Ganzes zu betrachten, könnte zwischen Maschinenbau, Elektrotechnik und weiteren ingenieurwissenschaftlichen Studiengängen unterschieden werden). Dies würde größere Stichproben erfordern im Vergleich zum ersten Zugang. Dafür erhielte man aber differenziertere Aussagen. Zu beachten ist, dass diese weiterhin auf Länderebene erfolgen würden, ohne Aussagen zur Leistung einzelner Universitäten oder Fachhochschulen zu liefern.

Eine solche AHELO-Konzeption, die in der Tradition von TEDS-M, TIMSS oder PISA stünde, ist von der Mehrheit der Länder, die an AHELO teilnehmen möchten, allerdings nicht gewollt, wie man hört. Es scheint eher um eine Evaluation der Studiengänge an einzelnen Hochschulen zu gehen. Damit stellt sich die Frage, was genau die Aussage sein soll, die am Ende möglich sein soll. Soll die Effektivität einzelner Hochschulen auf den Prüfstand gestellt werden? Was soll dann das Kriterium darstellen, anhand dessen man vergleichen will? Denkbar wäre in diesem Modell zum Beispiel, die am Ende der Ausbildung im Maschinenbau-Studiengang an der TU Darmstadt erreichten Kompetenzen mit denen im Maschinenbau-Studiengang an der Rheinischen Fachhochschule in Köln, denen im Studiengang Mechanical Engineering an der Purdue University in den USA und denen im Studiengang Mechanical Engineering an der University of the Punjab, Lahore (Pakistan) zu vergleichen.

Dieses Modell wäre prinzipiell denkbar. Eine Voraussetzung wäre, dass sich genügend Hochschulen bereit erklären teilzunehmen und dass sich ein Kernbestand an Kompetenzen identifizieren ließe, der der Ausbildung an allen vier Einrichtungen gerecht würde. Allerdings sind die noch existierenden Probleme nicht gering zu schätzen. Da institutionenbezogen getestet würde, würde auf eine vergleichsweise geringe Studierendenzahl zurückgegriffen werden müssen, die in ihrem Leistungspotenzial jahrgangsweise vermutlich starken Schwankungen unterlegen ist. Das bedeutet, dass sich einige Jahre später oder früher vermutlich andere Ergebnisse eines solchen Leistungsvergleichs ergeben würden.

In jedem Fall muss die am Ende einer Studie gewünschte Aussage vorab geklärt werden, da sie erhebliche Auswirkungen auf die Stichprobenziehung und die Ausgestaltung der Tests hat. Im

Zuge von TEDS-M wurde von einigen Teilnahmeländern beispielsweise nachträglich gewünscht, differenzierte Aussagen zu verschiedenen Typen an Lehrerausbildungen zu erhalten. Die Stichprobenziehung war von der IEA Jahre zuvor aber auf der nationalen Ebene angelegt worden und hatte dieses Merkmal „Programmtyp“ nur als Stratum berücksichtigt. Die resultierenden Zellenbesetzungen waren dementsprechend meist klein, der Stichprobenfehler entsprechend groß und eigenständige Vergleiche kaum möglich.

Kumulativität von Bildungsprozessen

Studien wie TIMSS oder PISA sind typischerweise als bilanzierende Studien am Ende von wichtigen Bildungsabschnitten konzipiert. Damit sagen sie etwas aus über die Gesamteffektivität von Bildungssystemen. Aussagen über die spezifische Leistung eines Bildungsabschnitts sind in der Regel nicht möglich, da kein Maß für die Eingangsvoraussetzungen der Auszubildenden zur Verfügung steht. Stellt dies für das Systemmonitoring ein eher geringes Problem dar, da es um übergreifende Effektivitätsfragen geht, wird diese Begrenzung bei institutionellen Evaluationen zu einer Falle.

Das Leistungsvermögen von Studierenden unterschiedlicher Institutionen unterscheidet sich bereits zu Beginn der Ausbildung gravierend. Eine Hochschule mit einem eher schwachen Bewerberpool – aus welchen Gründen auch immer, z. B. weil sie nicht in einer attraktiven Großstadt liegt oder weil sie in den handelsüblichen Rankings niedrig positioniert ist – hat weniger Möglichkeiten, leistungsstarke Studierende auszuwählen als eine Universität, die unter Tausenden von Bewerbungen auswählen kann, weil sie z. B. in Berlin oder München beheimatet ist oder weil sie in den Rankings weit oben positioniert ist. Nun kann die erstgenannte Hochschule in der Lehre eventuell deutlich besser arbeiten als die Letztgenannte. Aufgrund der Vermischung der Effekte von Lerngelegenheiten, die in der Ausbildung geboten wurden, und Eingangsselektivität im Endergebnis kann sie dieses ohne eine Kontrolle der Eingangsvoraussetzungen aber nicht zeigen, da sie im Zuge der Ausbildung zunächst einmal den Rekrutierungsvorteil der anderen Hochschulen aufholen muss.

Um die Lehreffektivität der verschiedenen Hochschulen angemessen darzustellen, müsste fairerweise also zunächst ein Maß gefunden werden, das die Unterschiede zu Beginn des Studiums anzeigt, sodass sie bei der Bilanzierung kontrolliert werden können. Im nationalen Vergleich könnte dies – trotz aller Probleme, die damit verbunden sind – die Abiturnote sein. Im internationalen Vergleich existiert dafür meines Wissens kein angemessenes Maß, sodass im Prinzip eine eigene Testung stattfinden müsste. Ein solches längsschnittliches Studiendesign wird in KoKoHs in vielen Projekten für einen kleinen Pool an Hochschulen angestrebt, ist in großem Maßstab aber schon national kaum denkbar, geschweige denn international.

Praktikabilitätsprobleme

In der Pilotierung wurden in AHELO einzelne Studierende aus der Liste der Eingeschriebenen zufällig gezogen und diese wurden dann getestet. Dass ein solches Vorgehen in sehr geringen Teilnahmequoten resultiert, kann sich jeder vorstellen, der die Realität der Hochschulen kennt. Die durch Prüfungsleistungen belasteten Studierenden kommen schlicht nicht zu zusätzlich anberaumten Testsitzungen. Anders als SchülerInnen können sie als Erwachsene auch nicht zu einer Teilnahme verpflichtet werden. Die üblichen Anforderungen an die Teilnahmequote, um Verzerrungen zu vermeiden (z. B. eine Teilnahmequote von 85%), sind also kaum zu erreichen. Damit stellt sich die Frage, wie solche Verzerrungen kontrolliert werden können (zum Beispiel anhand der bisheri-

gen Studienleistungen – dies würde aber eine weitreichende Kooperation der Hochschulverwaltungen aller teilnehmenden Institutionen erfordern, was kaum zu erwarten ist).

In TEDS-M wurden anders vorgegangen, indem Session Groups zufällig gezogen wurden. Das heißt dass aus der Liste an Lehrveranstaltungen eines Semesters für die avisierte Zielpopulation eine vorab definierte Zahl an Veranstaltungen zufällig gezogen wurde und in dieser dann im Rahmen der regulären Lehre die Testung stattfand. Dieses Verfahren weist ebenfalls Probleme auf (z. B. ein erhöhter Stichprobenfehler aufgrund der Clusterung der Studierenden), funktionierte aber besser als das AHELO-Verfahren.

Dennoch bleiben Herausforderungen, die nur mit voller Unterstützung der Hochschulleitungen und Anreizsysteme beispielsweise in Form von Studienpunkten oder finanzieller Entschädigungen gelöst werden können. In der Phase am Ende des Studiums besteht üblicherweise keine Anwesenheitspflicht. Zudem ist die Zahl der zu besuchenden Veranstaltungen gering, diese sind klein (15-20 Teilnehmende) und streuen stark. Die Definition der Zielpopulation ist ebenfalls schwierig: An welchem Kriterium kann in Deutschland, wo zumindest an den Universitäten das Studium nicht jahrgangswise organisiert ist, das letzte Jahr der Ausbildung festgemacht werden? Zudem resultieren die Anforderungen an die Stichprobengröße angesichts der geringen Studierendenzahlen pro Studiengang meist in eine Vollerhebung, die mit neuen logistischen Anforderungen verknüpft ist. Um diese Probleme zu lösen ist es fast zwingend, auch über die Option von Online-Testungen nachzudenken (trotz der damit neu entstehenden Probleme) und sich vielleicht stärker auf die Frage der Kontrolle möglicher Verzerrungen durch Selbstselektivitäten (z. B. in Form von Gewichtungungen) zu konzentrieren, als auf unerreichbaren Teilnahmequoten zu beharren.

Fazit

Stellt man die Chancen von Studien wie AHELO, TEDS-M oder KoKoHs ihren Problemen gegenüber, sollte das Fazit dennoch positiv ausfallen. Probleme sind schließlich dafür da, gelöst zu werden. Die Forschungsperspektiven, die mit der Arbeit an Theorie, Design und Instrumenten verbunden sind, sind äußerst reizvoll. Die Hochschulforschung stellt im Hinblick auf die direkte Kompetenzmessung eine der wenigen verbliebenen weißen Flecken dar. Allerdings muss präzise geklärt werden, welche Forschungsziele erreicht werden sollen. Die internationale Einbindung von AHELO böte die Chance durchzustarten und die im Zuge der Pilotierung gewonnene Expertise für eine qualitativ anspruchsvolle Studie zu nutzen. Der „Stein der Weisen“ wird AHELO voraussichtlich dennoch nicht sein. Dafür sind die zu bewältigenden Aufgaben zu umfangreich. Allerdings könnte mit einer Teilnahme Deutschlands national eine Dynamik entstehen, die zu weiteren Forschungen führt, wie dies bei den Schulleistungsstudien der Fall war. Die wichtigsten Forschungen entstanden hier erst nach den Large-Scale-Assessments durch nationale Ergänzungsstudien und eine Fülle an Mikro-Level-Analysen.

Eine entsprechende Forschungsdynamik könnte auch AHELO auslösen. Mit TEDS-M und KoKoHs wurde im Bereich der Messung akademisch erworbener Kompetenzen zwar eine erste Basis gelegt. Um mit den Anstrengungen anderer Länder in diesem Bereich mithalten zu können (siehe z. B. die *American College Testing- und das Assessing Learning in Australian Universities-Programme*) und nicht uneinholbar weit zurückzufallen, wäre es allerdings dringend geboten, sich hier stärker zu engagieren.

Literatur

- Baumert, J.; Kunter, M.; Blum, W.; Brunner, M.; Voss, T.; Jordan, A. et al. (2010).** Teachers' mathematical knowledge, cognitive activation in the classroom and student progress. *American Educational Research Journal*, 47(1), 133-180.
- Blömeke, S.; Suhl, U.; Döhrmann, M. (2012a).** Zusammenfügen was zusammengehört. Kompetenzprofile am Ende der Lehrerausbildung im internationalen Vergleich. *Zeitschrift für Pädagogik*, 58(4), 422-440.
- Blömeke, S.; Suhl, U.; Kaiser, G.; Döhrmann, M. (2012b).** Family background, entry selectivity and opportunities to learn: What matters in primary teacher education? An international comparison of fifteen countries. *Teaching and Teacher Education*, 28, 44-55.
- Blömeke, S., Zlatkin-Troitschanskaia, O.; Kuhn, C.; Fege, J. (Hrsg.) (2013).** Modeling and Measuring Competencies in Higher Education. Rotterdam: Sense Publishers.
- Braun, E. (2011).** Hochschulforschung in der Psychologie. *Die Hochschule. Journal für Wissenschaft und Bildung*, 2/2011, 74-90.
- Schmidt, W. H.; Blömeke, S.; Tatto, M. T. (2011).** *Teacher education matters. A study of the mathematics teacher preparation from six countries.* New York: Teacher College Press.
- Tatto, M.; Schwille, J.; Senk, S.; Ingvarson, L.; Peck, R.; Rowley, G. (2008).** *Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics: Conceptual framework.* East Lansing, MI: College of Education, Michigan State University.
- Tatto, M. T.; Schwille, J.; Senk, Sh. L.; Rodriguez, M.; Bankov, K.; Reckase, M. et al. (2012).** *The Mathematics Teacher Education and Development Study (TEDS-M). Policy, Practice, and Readiness to Teach Primary and Secondary Mathematics: First Findings.* Amsterdam: IEA.
- Zlatkin-Troitschanskaia, O; Kuhn, C. (2010).** *Messung akademisch vermittelter Fertigkeiten und Kenntnisse von Studierenden bzw. Hochschulabsolventen – Analyse zum Forschungsstand.* Johannes Gutenberg-Universität Mainz: Arbeitspapiere Wirtschaftspädagogik, 56.
- Zlatkin-Troitschanskaia, O.; Blömeke, S.; Kuhn, Ch; Buchholtz, Ch. (2012).** Wirksamkeitsprüfungen im Hochschulbereich – Aufgaben und Herausforderungen des BMBF-Forschungsprogramms „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“. *Zeitschrift für Evaluation*, 11(1), 95-103.

Outcome-Orientierung bei internationalem Vergleich von Studiengängen: Was kann AHELO von Tuning lernen?

Margarete Schermutzki & Robert Wagenaar

Tuning, die Rolle im Bologna Prozess und die internationale Entwicklung

Tuning Educational Structures¹ in Europe startete 2001 als ein Projekt europäischer Hochschulen, dass das bildungspolitische Ziel des Bologna-Prozesses, die Kompetenzorientierung von Studiengängen, begleiten und die Beschäftigungsfähigkeit der Studierenden fördern sollte. Im Tuning Projekt kommt der Forderung, dass Hochschulbildung neben der wissenschaftlichen Befähigung, auch auf die Befähigung in einem beruflichen Tätigkeitsfeld zu arbeiten (Employability), zur Persönlichkeitsbildung und zur Teilhabe am gesellschaftlichen Leben beitragen soll, eine zentrale Bedeutung zu.

Tuning unterscheidet in generische und fachspezifische Kompetenzen. Generische Kompetenzen sind fächerübergreifende Kompetenzen, die generell für ein akademisches Abschlussniveau gelten. Sie umfassen instrumentelle, interpersonelle und systemische Kompetenzen wie z. B. die Fähigkeit zur Analyse und Synthese, zur Reflexion des eigenen Handelns, zur kompetenten Handlungsfähigkeit mit komplexen und neuartigen Problemstellungen, sowie den systematischen Einsatz von wissenschaftlichen Methoden. Fachspezifische Kompetenzen beziehen sich auf die für den akademischen Abschluss erforderlichen Qualifikationsanforderungen für eine fachliche Disziplin. Sie umfassen Wissen, Verstehen und Anwenden sowie Wissensvertiefung und Wissensverbreiterung und sind unerlässlich für ein Fach bzw. eine Fachrichtung.

Im Tuning-Projekt wurden umfangreiche Befragungen von Absolventen, Arbeitgebern, Studierenden und Hochschullehrern in Europa und international zur Relevanz und zur Umsetzung von generischen und fachspezifischen Kompetenzen durchgeführt. Die Fragen wurden einvernehmlich von den beteiligten Hochschulen (Hochschullehrer und Studierende) unter Einbeziehung von Absolventen und Arbeitgebern für die jeweiligen Disziplinen entwickelt. Die Konzeption der Fragen basierte auf den Anforderungen an die akademischen Berufsfelder (Rolle der Disziplin in der Gesellschaft, den Profilen der von den teilnehmenden Hochschulen angebotenen Studiengängen, Tätigkeitsfelder der Absolventen und Positionen, in denen die Absolventen tätig sind), den Anforderungen an die wissenschaftliche Befähigung, die Persönlichkeitsbildung und die Verantwortung in der Gesellschaft für einen bestimmten akademischen Level.

Das Ergebnis der Befragungen zeigt die Bedeutung der generischen Kompetenzen und dass ihre Bedeutung in einer sich wandelnden Gesellschaft immer mehr zunimmt.

Von den im Tuning beteiligten Hochschulen wurden aus den Ergebnissen der Befragung theoretisch fundiert unter Berücksichtigung von generischen und fachspezifischen Kompetenzen Disziplin-bezogene Referenzrahmen für eine große Anzahl von Bereichen entwickelt. Hierbei wurden zunächst die für eine Disziplin maßgeblichen fachspezifischen Kernelemente gemeinschaftlich festgelegt und die Referenzrahmen dann unter Einbeziehung von generischen Kompetenzen entwickelt.

¹ <http://www.unideusto.org/tuningeu/>

Die so entwickelten fachbezogenen Referenzrahmen im Sinne von Tuning sind als Instrumente für eine outcome- bzw. kompetenzorientierte Studiengangentwicklung zu verstehen. Hierbei sind generische und fachspezifische Kompetenzen sinnvoll in einem Studiengang zu verbinden.

Die Anwendung der durch das Tuningprojekt entwickelten fachbezogenen Referenzrahmen schafft Transparenz, Vergleichbarkeit (national und international) und Kompatibilität. So kann Hochschulen die Möglichkeit geboten werden, ihre Studiengänge individuell unter Berücksichtigung von unterschiedlichen regionalen und kulturellen Besonderheiten zu entwickeln und durch Benchmarking international, national, regional, fachbezogen zu vergleichen sowie ihr Profil zu schärfen und die Qualität der Studiengänge zu sichern und weiter zu entwickeln. Durch die Anwendung von Tuning-Referenzrahmen werden Möglichkeiten für die Anerkennung von Leistungen geschaffen und gleichzeitig die wesentlichen Unterschiede der einzelnen Studiengänge erhalten.

Bei der Entwicklung der fachbezogenen Referenzrahmen kommt der Ergebnisorientierung eine hohe Bedeutung zu, um die Absolventen auch auf ihre zukünftige Rolle in Beruf und Gesellschaft vorzubereiten. Entsprechend rücken die zu erwerbenden Kompetenzen in den Vordergrund. Die Prüfbarkeit von Kompetenzen erfordert die Definition von Lernergebnissen. Der Vorteil liegt in der Verständlichkeit, Transparenz und Prüfbarkeit dieser sprachlichen Lösung.

Die Darstellung von vergleichbaren Lernergebnissen in den entsprechenden Qualifikationsstrukturen des europäischen und des deutschen Qualifikationsrahmens für Hochschulabschlüsse und des Europäischen und des deutschen Qualifikationsrahmens für Lebenslanges Lernen (Qualifikationsrahmens LLL) steht im Einklang mit den in Tuning entwickelten fachbezogenen Referenzpunkten.

Neben den fachbezogenen Referenzrahmen wurden von den im Tuning beteiligten Hochschulen ein Modell und eine Prozessbeschreibung zur Konzeption von Outcome-orientierten Studiengängen entwickelt. Es wurden Planungshilfen zur Formulierung von Lernergebnissen, ihrer Vermittlung und ihrer Prüfung, zur Planung und Erfassung der Arbeitsbelastung von Studierenden und zur Qualitätsentwicklung erstellt (González & Wagenaar, 2008; Lokhoff et al., 2010).

Um die Gestaltung kompetenzorientierter Studiengänge zu unterstützen, bietet die neu gegründete Tuning Academy² Möglichkeiten, Lehrende, Hochschulentwickler, Qualitätsmanager und Curriculumsentwickler zu unterstützen bei der kompetenzorientierten Curriculumsentwicklung, bei der Level bezogenen Definition von generischen und fachbezogenen Lernergebnissen, bei kompetenzförderlichen Lehr/Lernarrangements, kompetenzbasierten Prüfungen, Kompetenzentwicklungsständen von Studierenden, Maßnahmen zur Förderung von Kompetenzerwerb von Studierenden, kompetenzorientierter Evaluation, Qualitätsentwicklung und Förderung von Lehrenden im Hinblick auf die Kompetenzorientierung der Lehre.

Die fachbezogenen Referenzrahmen, Methoden und Instrumente von Tuning finden international Anerkennung und Anwendung. Es ist den beteiligten Hochschulen so gelungen, bei der Gestaltung des neuen europäischen Hochschulraumes aktiv mitzuarbeiten und Einfluss auf die Umsetzung des Bologna-Prozesses zu nehmen.

Kompetenzorientierung an deutschen Hochschulen

Gemäß Beschluss der KMK von 2000 sind neue Studiengänge in Deutschland zu akkreditieren. Grundlage und Maßstab für die Akkreditierung ist der „Qualifikationsrahmen für deutsche Hochschulabschlüsse“ (HQR). Der HQR orientiert sich an der Einteilung der Kategorien von erworbenen

² <http://www.unideusto.org/tuningeu/tuning-academy.htm>

Kompetenzen am TUNING-Projekt und an den von der Joint Quality Initiative entwickelten „Dublin Descriptors“. Die Empfehlungen des Tuning-Projektes zur kompetenzbasierten Studiengangentwicklung, unter Berücksichtigung von generischen und fachspezifischen Kompetenzen stehen somit in Einklang mit den Rahmenvorgaben und Empfehlungen der KMK und der HRK.

Grundsätzlich sind die Bundesländer für die gesetzliche Verankerung von Outcomeorientierung von Studiengängen sowie die Verwendung von Kompetenzen und Lernergebnissen an deutschen Hochschulen zuständig. Die gesetzliche Verankerung ist in den verschiedenen Bundesländern unterschiedlich gestaltet, aber häufig in der Auflage zur Akkreditierung von Studiengängen impliziert.

Hochschulen haben die Möglichkeit, Kompetenzen und Lernergebnisse strategisch im Rahmen von Zielvereinbarungen (mit den Ministerien und den Fachbereichen der Hochschulen) zu verankern. Eine Verankerung kann auch in den Rahmenprüfungsordnungen und Prüfungsordnungen erfolgen.

Die Umsetzung der Rahmenvorgaben der KMK und Empfehlungen zur Studiengangentwicklung der HRK gelingt allerdings nur unzureichend. Wie Akkreditierungsagenturen berichten (Jaudzims, 2011), ist durch die Einführung der Akkreditierung das Niveau der Outcome-Orientierung von Studiengängen gestiegen. Der Aufwand für die Überprüfung von Lernergebnissen in Modulbeschreibungen und ihr Bezug zu den entwickelten Studiengängen in den Akkreditierungsverfahren sind aber zu hoch und nicht im erforderlichen Umfang zu leisten.³

Probleme entstehen mit der unzureichenden Berücksichtigung von generischen Kompetenzen als „learning outcomes“ bzw. Qualifikationszielen von Studiengängen über wissensbasierte Kompetenzen hinaus. Diese Probleme werden verstärkt durch unzureichende Umsetzung von kompetenzbasierten Lehr-/Lern- und Prüfungsmethoden und Evaluationsmethoden die wenig Bezug auf die Lernergebnisse, Qualifikationsziele und die zu evaluierenden Module haben.

Eine Untersuchung von Modulbeschreibungen hat ergeben, dass Module vor allem aus Vermittlungsperspektive beschrieben werden. In vielen Fällen bleibt die zu erwerbende Kompetenz im Modul, die Art der Vermittlung und Prüfung unklar; Lehr-, Lern- und Prüfungsaktivitäten sind nicht abgestimmt mit der Lernsituation; Prüfungsformen erscheinen eher ungeeignet, um das Erreichen der Lernziele adäquat nachzuweisen. Employability und Citizenship erscheinen häufig als „leere Hüllen“. Darüber hinaus ergeben sich technische Probleme bei der Abbildung von Modulen in Form von Lernergebnissen. Es werden häufig Lehrveranstaltungen zu Modulen zusammengefasst und abgebildet und nicht Module als Einheit gesehen. Entsprechend werden von Studierenden Lehrveranstaltungen und nicht Module evaluiert.

Zusammenfassend ist zu sagen, dass der Perspektivwechsel von der Input- zur Outcome-Orientierung von Studiengängen insbesondere in den Modulen nicht ausreichend umgesetzt wurde. Veränderungen aus der Sicht der Lehrenden, verbunden mit didaktischer Weiterbildung, sind erforderlich.

Ein internationaler Vergleich von kompetenzbasierten Studiengängen an deutschen Hochschulen ist daher zum jetzigen Zeitpunkt nicht zielführend.

Es stellt sich zunächst die Frage, wie deutsche Hochschulen von der Notwendigkeit und vom Nutzen von kompetenzbasierten Studiengängen zu überzeugen sind:

Zunächst ist es wichtig, dass die Hochschulen die an den Studiengängen beteiligten Lehrenden in die Studiengangentwicklung einbeziehen. Gemeinsam und im Team werden zu erreichende Lernergebnisse des Studiengangs unter Berücksichtigung von fachbezogenen und generischen Kompetenzen entwickelt sowie die geeigneten Lehr-, Lern- und Prüfungsmethoden festgelegt.

³ Gespräch mit Akkreditierungsagentur

Die Entwicklung von kompetenzbasierten Studiengängen bietet deutschen Hochschulen die Möglichkeit, das Innovationspotential von deutschen Hochschulen, der Region oder der Hochschule zu verdeutlichen, wesentliche Unterschiede zu anderen Hochschulsystemen und Kulturen beizubehalten und ihr eigenes Profil zu schärfen. Die konsequente Umsetzung von kompetenzorientierten Studiengängen kann deutsche Hochschulen unterstützen, Motor zu sein für die dringend notwendigen Veränderungen und Anforderungen an die europäischen Hochschulen bezogen auf die internationale Konkurrenz.

An deutschen Hochschulen wird die kompetenzbasierte Studiengangentwicklung als Eingriff in die Freiheit der Lehre angesehen und hinter den von Tuning entwickelten fachspezifischen Referenzrahmen und Instrumenten wird eine Standardisierung von Studiengängen vermutet. Dabei werden wesentliche Aspekte wie das eigene Bedürfnis nach Vergleichbarkeit insbesondere in Bezug auf gute Forschung und Lehre ausgeblendet.

Auch die deutschen Hochschulen sind längst in eine grenzenlos globalisierte Welt geraten, in der Rankings eine immer wichtigere Rolle übernehmen und Hochschulen anhand von Indikatoren bewertet werden, die regionale und nationale Aspekte nicht berücksichtigen. Die Ausrichtung der Studiengänge auf generische und fachspezifische Kompetenzen bietet den deutschen Hochschulen die Möglichkeit, ihre Stärken im Bereich von Forschung, Innovation, Ideenreichtum vor Ort an den Hochschulen einzubringen und auch nach außen transparent zu machen.

Durch die Bereitschaft, sich mit internationalen Hochschulen zu vergleichen, können deutsche Hochschulen in dieser globalen Situation ihre lokale Identität stärken. Wir brauchen engagierte Hochschullehrerinnen und -lehrer, die den Willen zur Kooperation zwischen Wirtschaft, Forschung, Wissenschaft haben sowie die regionalen Anforderungen und die Anforderungen an die Absolventen kennen und die bereit sind, an der Entwicklung und Umsetzung des europäischen Hochschulraums als Antwort auf die Globalisierung mitzuarbeiten und diesen zu stärken.

Literatur

González, J.; Wagenaar, R. (2008). Tuning Educational Structures in Europe. Universities' contribution to the Bologna Process. An introduction, Bilbao 2008

Jaudzims, Susanne (2011). Learning Outcomes in der Hochschulbildung Chancen, Herausforderungen und Grenzen eines lernerorientierten Ansatzes in der Hochschullehre. Onlinedokument: <http://d-nb.info/1019362642/34>

Lokhoff, J.; Wegewijs, B.; Durkin, K.; Wagenaar, R.; González, J.; Isaacs, A.; Donà della Rose, L.; Gobby, M. (2010). Tuning Educational Structures in Europe, A Guide to Formulating Degree Programmes Profiles, Including Programmes Competences and Programme Learning Outcomes, Bilbao, Groningen and The Hague., 2010, ISBN 978-84-9830-375-9.

Qualifikationsrahmen für Deutsche Hochschulabschlüsse (Im Zusammenwirken von Hochschulrektorenkonferenz, Kultusministerkonferenz und Bundesministerium für Bildung und Forschung erarbeitet und von der Kultusministerkonferenz am 21.04.2005 beschlossen.)

Zur Anlage von AHELO

Vor- und Nachteile Internationaler Vergleiche im Kontext des AHELO-Projekts⁴

Dominic Orr

Einleitung

Die Diskussion um den Zweck und die Ziele des AHELO-Projekts führt zurück auf die Frage des Sinns und Zwecks von derartigen vergleichenden Studien im Bildungsbereich. Dieser Frage wurde im Rahmen eines Workshops bei der GfHf-Tagung gesondert nachgegangen. Das AHELO-Projekt hat zum Ziel, eine Messung von Lernergebnissen (learning outcomes) und – zumindest perspektivisch – von Lernzuwachs (learning gains) in Hochschulen zu bewerkstelligen. Handelnden Akteuren soll auf diesem Wege eine objektive Basis für Entscheidungen gegeben werden. Damit soll dieses Projekt Steuerungswissen zur Effektivität von Lehr- und Lernaktivitäten bereitstellen, das unmittelbar vom Hochschulleitungspersonal und der Hochschulpolitik verwendet werden und mittelbar die Entscheidungen von zukünftigen Studierenden zur Studienaufnahme beeinflussen kann. Vor diesem Hintergrund diskutierten die Teilnehmer(innen) des Workshops die Vor- und Nachteile international vergleichender Hochschulforschung, um daraus die diesbezüglichen Herausforderungen für AHELO zu formulieren.

Vor- und Nachteile internationaler Vergleiche

Tabelle 1: Vor- und Nachteile laut Workshopteilnehmer(innen)

Vorteile
<ul style="list-style-type: none"> ■ Suche nach funktionalen Äquivalenten (Erkenntnis über Funktionen) ■ Lerneffekt durch Infragestellung der eigenen Position/von Selbstverständlichkeiten (Lernprozess) ■ Bewusstsein über eigene Besonderheiten setzt Differenzenerfahrung voraus ■ Möglichkeit der systematischen Untersuchung des Einflusses von Kontextunterschieden (Varianz auf Makroebene) ■ Internationale Einordnung: Besonderer Fall oder Regelfall?
Nachteile
<ul style="list-style-type: none"> ■ Aufwendig ■ Gefahr der mangelnden Berücksichtigung des Kontexts ■ Gefahr der Unschärfe von Kategorien ■ Mangelnde Vergleichbarkeit

Die diskutierten Vor- und Nachteile sind in der Tabelle 1 wiedergegeben. Bei den Nachteilen handelt es sich um die Problematik der Definition und Abgrenzung eines geeigneten Tertium comparationis. Bei einem Vergleich von funktionalen Äquivalenten muss zunächst das Gemeinsame bestimmt werden, in dem zwei verschiedene Gegenstände oder Sachverhalte übereinstimmen. Diese Arbeit ist sehr aufwendig, weil sie konzeptionelle Entwicklungen erfordert. Beispielsweise

⁴ Der Autor dankt Kai Mühleck für seinen Beitrag zum Workshop und zu diesem Text.

heißt die Qualifikation zum Abschluss eines zur Studienberechtigung führenden Bildungsgangs in der schulischen Oberstufe in Deutschland „Abitur“. Das Abitur gibt es aber nur in Deutschland. Daher müssen Kriterien bestimmt werden, die ähnliche Abschlussprüfungen erkennen lassen. In vielen Ländern gibt es Abschlüsse, die die Hochschulreife eines Schulabgängers bescheinigen. In Österreich und Polen heißt diese Qualifikation „Matura“, in Italien „Maturità“, in Frankreich „Baccalauréat“ und in Spanien „Bachillerato“. Wie die Namen bereits implizieren, haben alle Qualifikationen etwas mit der akademischen Reife zu tun. Sie unterscheiden sich aber sehr in der jeweiligen Einbettung der Qualifikationen in den nationalen Bildungssystemen (Orr/Riechers, 2010). Wenn also die Anzahl der erfolgreich bestandenen Abschlussprüfungen für die schulische Oberstufe verglichen wird, grenzt eine mangelnde Berücksichtigung des Kontexts (in diesem Fall: der Qualifikationsstrukturen eines Bildungssystems) den Wert des Vergleichs ein.

Die Vorteile von Vergleichen für die Bildungsforschung ergeben sich aber wiederum direkt aus der Suche nach funktionalen Äquivalenten (Phillips, 2006). Dabei werden die Funktionen etwa der Abiturprüfung konkretisiert und können neu reflektiert werden. Durch diesen Prozess eröffnet die vergleichende Hochschulforschung die Perspektive auf Alternativen, die etwa die gleichen oder ähnliche Funktionen erfüllen. Es können (national)systemimmanente „Selbstverständlichkeiten“ und damit existierende Annahmen, organisatorische Verfahren sowie Selektionsmechanismen in Frage gestellt werden.

Die vergleichende Hochschulforschung ist darüber hinaus in der Lage, den komplexen Prozessen der Globalisierung Rechnung zu tragen. Im Kontext der gleichzeitigen Angleichung und Ausdifferenzierung (Schriewer, 2007) zwischen Hochschulen und Hochschulsystemen ist eine wichtige Frage, ob das in einem (nationalen) Kontext existierende Phänomen eine Besonderheit oder den Regelfall in vielen nationalen und lokalen Kontexten darstellt.

Besondere Herausforderungen des Vergleichs von Lehr- und Lernleistungen im Hochschulbereich

Das AHELO-Projekt will die Kompetenzen von Bachelorstudierenden kurz vor ihrem Abschluss messen, um (i) ein objektives Bild der Lernergebnisse zu ermöglichen, die an (ii) unterschiedlichen Hochschulen in (iii) unterschiedlichen Ländern erzielt werden. Alle drei Aspekte stellen zusammen eine besondere Herausforderung für den Vergleich sowie für die Hochschulforschung dar.

Da AHELO Lernergebnisse untersuchen will, muss die folgende Frage beantwortet werden: Können die beabsichtigten Lernergebnisse in den Hochschulen unterschiedlicher Länder so definiert werden, dass ein Vergleich möglich wird? Zur Lösung der Herausforderung einer Definition vergleichbarer und international akzeptierter Lernergebnisse greift man im AHELO-Projekt auf die Arbeiten des TUNING-Projekts zurück. Dieses Projekt hat auf hochschulischer Ebene die standardisierte Formulierung von Lernergebnissen und Kompetenzen zum Ziel, die durch einen spezifischen Studiengang vermittelt werden sollen. Hochschulen aus allen fünf Kontinenten nehmen mittlerweile an diesem Projekt teil.⁵ Auf der Basis dieser Arbeit ist es möglich, über die induktive Herleitung und Expertendiskussionen auch die allgemein erwarteten Lernergebnisse und Kompetenzen für einen Fachbereich zu definieren. Die TUNING-Verantwortlichen haben dies im Rahmen von AHELO für die Fachbereiche Ingenieurwissenschaften und Wirtschaftswissenschaften durchgeführt.⁶ Diese Methode ist geeignet, eine Basis für die Kompetenzmessung im Ländervergleich zu definieren. Dennoch handelt es sich dabei immer noch um normative Ziele und nicht

5 <http://www.unideusto.org/tuningeu/>

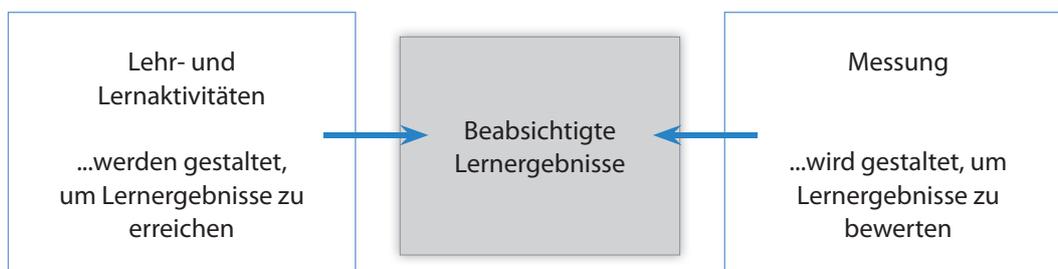
6 <http://www.oecd.org/edu/highereducationandadultlearning/ahelodocuments.htm>

um konkrete Items, die getestet werden. Diese wurden anhand der normativen Ziele definiert. Nach dem „Übersetzungsprozess“ von normativen Zielen hin zu Testing-Items wurden die Items in Fokusgruppen und mittels Pre-testing auf ihre Messqualität geprüft. Laut dem ersten Machbarkeitsbericht fiel diese Evaluation positiv aus, obwohl nicht alle normativen Vorgaben zu den erwarteten Kompetenzen in den Tests realisiert werden konnten (OECD 2013: 118f). Ähnlich ist man bei dem Fachbereich Ingenieurwesen vorgegangen.

Hier wurde ein nicht unumstrittenes Instrument zur Kompetenzmessung aus den USA adaptiert, das Collegiate Learning Assessment (CLA) (Douglass et al., 2012). Die Akzeptanz dieses Instrument ist viel geringer (OECD 2013: 62); u. a. wird kritisiert, dass dieses Instrument nur für den amerikanischen Hochschulsektor geeignet ist. Die Kritik hebt hervor, dass über das Instrument indirekt das amerikanische System der Hochschulbildung und Kompetenzmessung den Hochschulen anderer Länder „aufgezwungen“ würde (OECD 2013: 62). In der englischen Literatur werden solche Effekte unter dem Begriff „backwash“ diskutiert (Biggs, 1996). „Backwash“ beschreibt den Einfluss von Tests auf den eigentlichen Inhalt von Lehr- und Lernaktivitäten.

Biggs stellt Lehr- und Lernaktivitäten einerseits und Messung andererseits einander gegenüber, um das Problem der „konstruktiven Angleichung“ zu thematisieren (siehe Abbildung 1). Eine Veränderung in Lehr- und Lernaktivitäten kann zur Verbesserung der gemessenen Ergebnisse führen und dies stellt auch das Ziel der Messung dar. Die Herstellung eines solchen Zusammenhangs wird in der didaktischen Forschung jedoch deshalb kritisch betrachtet, da die Wirkweise in beide Richtungen verlaufen kann. Das heißt, die Hochschullehrer(innen) und die Studierenden fangen an, ihren Unterricht bzw. ihre Lernstrategie nach den Tests zu richten („teaching to the test“). Dies konnte beispielsweise für Lehrevaluationen im deutschen Raum nachgewiesen werden (Nowakowski et al., 2012).

Abbildung 1: „constructive alignment“ nach Biggs



Quelle: nach Biggs (1996), eigene Darstellung

Für eine international vergleichende Studie, die vielen Restriktionen unterliegt, um einen Vergleich möglich zu machen, hat die angesprochene mögliche direkte Rückwirkung der Messung auf Lehr- und Lernaktivitäten zur Folge, dass ein hoher Anspruch an die Messinstrumente zu stellen ist; dies kann eventuell sogar zur Überstrapazierung führen. Im ersten Machbarkeitsbericht wird bereits von diesem Effekt im Fallstudien-Land Kanada (Ontario) berichtet (OECD 2013: 193). Hier wird allerdings die Kompetenzorientierung und nicht die Test-Items angesprochen.

„Faculty members were further encouraged by the potential of AHELO when a small group came together to score the constructed responses of the students. The framing of questions to make students ‘think like an engineer’ was innovative to some – prompting them to consi-

der how they taught their students, what they expected of them and how they were assessed. "I wish I had considered these issues before setting the final exam for my students," noted one professor during the scoring session, "I would have done it much differently."

Wenn jedoch die Kompetenzmessung keinen Einfluss auf die Didaktik der Hochschullehre hätte, wäre die Zielrichtung des AHELO-Projekts zu eingeschränkt und es ließe sich fragen, ob AHELO nur ein neues Hochschulranking sein will. Im Falle einer ausschließlichen Messung von Kompetenzen, um also einen neuen Indikator für die Messung von Hochschulperformanz anzubieten, wäre der Aufwand als zu hoch einzuschätzen. Besonders deshalb, weil die relative Performanz zwischen Hochschulen, die mittels Kompetenzmessung am Ende des Studiums wiedergegeben werden soll, sich vermutlich in vielen Fällen auch über Variablen für die schulische Leistung von Studierenden vor ihrem Studium „vorhersagen“ lässt; d. h. es wird nicht geprüft inwiefern die Hochschulen einen eigenen Beitrag zur Kompetenzsteigerung gemacht haben. Deshalb sind zwei zusätzliche Aspekte im AHELO-Instrumentarium vorgesehen: (i) Kontextinformationen und (ii) die Messung von Lernzuwachs (learning gains).

Im Projekt werden Kontextinformationen über Befragungen von Studierenden, Wissenschaftler(inne)n und über Standardindikatoren aus dem Hochschulvergleich (U-Multirank) erhoben. Diese sollen eine differenzierte Betrachtung der Ergebnisse der Kompetenzmessungen ermöglichen. Sie sollen laut Projektbericht einen „quality space“ im gesamten Datensatz ermöglichen, der Profile von Hochschulen anhand mehrerer Kriterien (nicht nur Kompetenzmessungen) besser erkennen lässt (OECD 2013: 67, 91, 94).

Die Messung von Lernzuwachs ist darüber hinaus ein wichtiges Alleinstellungsmerkmal des AHELO-Projekts, das damit die Leistungsmessung von Hochschulen wirklich auf hochschulische Leistungen beziehen kann. Nur dieser Aspekt – zusammen mit der Messung und der Berücksichtigung von Kontextinformationen – lässt einen klaren Rückschluss auf die Lehr- und Lernaktivitäten sowie die Lernerfahrungen von Studierenden in verschiedenen Hochschulen zu. Dieser Aspekt wurde bisher überhaupt nicht bei globalen Hochschulvergleichen untersucht und würde daher den Aufwand des Projekts insgesamt rechtfertigen. Obwohl dies ein Ziel des Projekts ist, hat man sich während der Machbarkeitsstudie entschieden, den Versuch des Vorher-Nachher-Vergleichs nicht zu unternehmen (OECD 2013: 80). Vielmehr trifft sich eine Arbeitsgruppe Anfang 2013, um die Möglichkeiten eines solchen Vergleichs zu erörtern.

Diskussion

Die Debatten im Workshop schlossen mit einer in der Gruppe erarbeiteten Gegenüberstellung von möglichen Problemen und diesbezüglichen Lösungsvorschlägen ab. Diese werden in Tabelle 2 wiedergegeben.

Der erste Diskussionspunkt betraf den Aufwand, der für die Realisierung des Projekts notwendig ist. Aufgrund seiner Zielsetzungen kann man das AHELO-Projekt mit den drei ebenfalls international-vergleichenden Studien U-Multirank, PISA und PIAAC vergleichen. Bei U-Multirank geht es darum, einen multidimensionalen Vergleich zwischen Hochschulen zu ermöglichen, der die Differenzen zwischen Hochschulprofilen explizit berücksichtigt – was bei den meisten Hochschulrankings bislang nicht geschieht. Auch das AHELO-Projekt hat dieses Ziel, es geht aber darüber hinaus, wenn es die Hochschulperformanz über das Konstrukt Kompetenzerwerb vergleicht. Bei PISA und PIAAC will man die Kompetenzen von Personen in verschiedenen Ländern im Schulalter (PISA) sowie im Erwachsenenalter (PIAAC) vergleichen. Auch das AHELO-Projekt hat dies vor; al-

Tabelle 2: Diskussionspunkte und erste Lösungsvorschläge der Workshopteilnehmer(innen)

	Probleme/Diskussionspunkte	„Lösungsvorschläge“
1	Brauchen wir AHELO noch (wenn wir bereits PISA, PIAAC und Multirank haben)?	Das Alleinstellungsmerkmal ist der Aspekt „Learning gains“.
2	Negative Effekte auf Lehr- und Lernaktivitäten durch den sogenannten Backwash-Effekt	Tests müssen einen didaktischen Anspruch haben (Tuning-Projekt). Es gibt einen trade-off zwischen einem guten Test und der Pflege von guten Lehr- und Lernaktivitäten.
3	Der jeweilige nationale Kontext muss bei Frageformulierungen und Antwortkategorien berücksichtigt werden.	Dies setzt starke nationale Teams voraus, die den trade-off zwischen Kontextualisierung und Vergleichbarkeit leisten müssen.
4	Der institutionelle, kulturelle und regulative Kontext einer Hochschule in einem Land muss auch bei der Interpretation der Ergebnisse berücksichtigt werden.	Dies ist nicht prinzipiell problematisch, stellt aber methodische Herausforderungen dar. Gleichwohl wird es v. a. die nationale Forschung sein, die Unterschiede mittels tiefergehender Studien erklären kann (ähnlich wie bei PISA).

lerdings fokussiert das Projekt auf Erwachsene in der Hochschulbildung und hat im Unterschied zu den anderen zwei Projekten eine Bildungsinstitution (die Hochschule) und nicht ein Land als Vergleichsobjekt. In der Diskussion ging es jedoch darum, ob der Aufwand sich für AHELO wirklich lohnt, wenn es ähnliche Projekte gibt, die Teilaspekte bereits abdecken. Die Gruppe erkannte jedoch an, dass die Messung von Lernzuwachs innerhalb der Hochschule unter Berücksichtigung von Kontextfaktoren wirklich etwas Innovatives darstellt. Leider wurde diese Komponente in der aktuellen Machbarkeitsstudie noch nicht voll ausgearbeitet. Zum gegenwärtigen Zeitpunkt – d. h. vor dem Abschlussbericht der Machbarkeitsstudie) weiß man über die Pläne der OECD hierzu zu wenig.

Die Effekte von Testing auf Lehr- und Lernaktivitäten wurden ebenfalls von der Gruppe diskutiert. Die Teilnehmer(innen) am Workshop hoben hervor, dass es wichtig sei, die Effekte kritisch zu betrachten. Die Test-Instrumente müssten einen didaktischen Anspruch erfüllen, was aufgrund der Erfahrungen des TUNING-Projekts bei den fachspezifischen Tests relativ gut gelingt. Die normativen Ziele für die fachspezifischen Tests werden im Rahmen des TUNING-Projekts auf hochschulischer Ebene entwickelt, dann verallgemeinert und schließlich in Test-Items „übersetzt“. Bei den generischen Kompetenzen ist es sicher schwieriger, allgemeine normative Ziele zu finden, die auch unabhängig von nationalen und hochschulischen Kontexten gelten können. In dieser Domäne schlagen sich die Nachteile von internationalen Vergleichen, wie Unschärfe von Kategorien und mangelnde Vergleichbarkeit zwischen Kompetenzmessungen in verschiedenen Ländern, in Hochschulen mit unterschiedlichen Profilen und in unterschiedlichen Fachbereichen, besonders stark nieder. Engagierte nationale Teams sind eine Voraussetzung zur Unterstützung dieses Prozesses der Angleichung, aber es bleibt fraglich, ob das Ziel innerhalb des AHELO-Projekts erreichbar ist.

Das Problem der adäquaten Berücksichtigung von institutionellen, kulturellen und regulativen Kontexten wird zwar direkt im Projekt angegangen. Dennoch blieb es ein Diskussionspunkt zwischen den Teilnehmer(inne)n, denn eine volle Berücksichtigung dieser Kontextfaktoren – die die

funktionalen Äquivalente faktisch in Frage stellt – macht den Vergleich letzten Endes schwierig. Ein internationaler Vergleich schränkt immer die Berücksichtigung von besonderen Eigenschaften ein. Allerdings sieht man schon am Beispiel von PISA, dass einer der positiven Effekte solcher Studien darin liegt, nationale – tiefergehende – Studien voranzutreiben. Daher bleibt es bei solchen Projekten eine Aufgabe der nationalen Forschung, sowohl die Weiterentwicklung der Forschungsmethoden und der dazu gehörigen methodischen Konstrukte zu unterstützen als auch die Ergebnisse der Studien als Ausgangspunkt für tiefergehende Studien zu verwenden. Ohne diese Schritte wird AHELO zwar ein interessantes Benchmarking-Instrument bereitstellen, aber sein Beitrag zum Steuerungswissen im Hochschulsystem wird gering bleiben.

Literatur

- Biggs, J. (1996).** Aligning teaching for constructing learning. The Higher Education Academy. Online verfügbar unter: http://www.heacademy.ac.uk/assets/documents/resources/database/id477_aligning_teaching_for_constructing_learning.pdf
- Douglass, J. A.; Thomson, G.; Zhao, C.-M. (2012).** The learning outcomes race: the value of self-reported gains in large research universities. In: *Higher Education*, 64: 317–335.
- Nowakowski, A.; Vervecken, D.; Braun, E.; Hannover, B. (2012).** Der differentielle Einfluss prozess- versus ergebnisorientierten Lehrevaluations-Feedbacks auf die Lehrorientierungen von Hochschuldozierenden. *Zeitschrift für Erziehungswissenschaft*, 15, 253-271.
- OECD (2013).** AheLO Feasibility Study – Volume 1. Online verfügbar unter: <http://www.oecd.org/edu/highereducationandadultlearning/AHELOFSReportVolume1.pdf>
- Orr, D.; Riechers, M. (2010).** Organisation des Hochschulzugangs im Vergleich von sieben europäischen Ländern. HIS:Forum Hochschule 11/2010. Online verfügbar unter: http://www.his.de/pdf/pub_fh/fh-201011.pdf
- Phillips, D. (2006).** Comparative education: method. *Research in Comparative and International Education*, 1(4): 304-319.
- Schriewer, J. (Hg. 2007).** Weltkultur und kulturelle Bedeutungswelten – Zur Globalisierung von Bildungsdiskursen. Frankfurt: Campus Wissenschaft.

Chancen und Grenzen von Rankings: Stand der Forschung und Bezug zu AHELO

Barbara M. Kehm

Das Konzept des Workshops sah eine Gliederung in zwei Teile vor: Zunächst wurden in einem Impulsreferat wesentliche Chancen und Grenzen von Rankings im Allgemeinen vorgestellt und argumentativ untermauert. Sodann erfolgte eine Überleitung zum zweiten Teil, in dem die Teilnehmerinnen und Teilnehmer die Möglichkeiten und Schwierigkeiten von AHELO vor dem Hintergrund dieser Einstiegspräsentation diskutierten (s. a. den Text von Susanne In der Smitten in diesem Band).

Als grundlegende Chancen und Grenzen von Rankings können folgende Aspekte thesenartig herausgestellt werden:

- (1) Die meisten Rankings aggregieren Daten zu einer Reihe unterschiedlicher, meist auf die Forschung bezogener Indikatoren, auf deren Grundlage dann ein institutioneller (ggf. auch fachbezogener) Listenplatz zugewiesen wird. Reputation spielt dabei oft eine größere Rolle als Performanz. Das so dargestellte Wissen spiegelt die Leistung des Fachbereichs oder einer Institution nicht angemessen wider und sagt kaum etwas über die Qualität der Lehre aus.
- (2) Es gibt kein objektives Ranking (vgl. Hazelkorn, 2011): Die Wahl der Indikatoren und ihre Gewichtung enthalten zwangsläufig Werturteile. Die Präferenzen der Ranking-Produzenten sowie die Zugänglichkeit von Daten haben einen großen Einfluss auf die Ranking-Ergebnisse. Insofern zählen Rankings eher das, was gemessen werden kann, als sie messen, was zählt (vgl. Locke, 2011). Der Gebrauch von Proxy Variablen kann die Ergebnisse zusätzlich verzerren, weil häufig keine logische Beziehung zu Aussagen über Qualität und Performanz besteht. So wird z. B. der Abiturnotendurchschnitt als Indikator für die Qualität sich neu einschreibender Studierender genutzt, sagt aber nichts über die Qualität der Studienerfahrung aus. Ähnlich ist es bei dem Indikator ‚Selektivität eines Studiengangs‘, der natürlich nichts über die Qualität des Studiengangs als solche aussagt (vgl. Hazelkorn, 2011: 51). Auch liegt Rankings meist eine enge Definition von Qualität zugrunde, die dann auf alle untersuchten Einrichtungen gleichermaßen angewendet wird. Dies führt zu so genannten One-fits-all-Messungen, bei denen etwa 95% aller Hochschulen weltweit an Kriterien gemessen werden, die nur für die oberen 5% der Forschungsuniversitäten angemessen erscheinen. Durch Rankings bekommt der Status einer Hochschule einen höheren Wert als das, was die Studierenden lernen bzw. gelernt haben.
- (3) Akademische wie kommerzielle Rankings beanspruchen vielfach, Informationen für die Hochschul- und Studienfachwahl bereitzustellen. Es stellt sich jedoch die Frage, wie viele Studieninteressierte tatsächlich nach hoch gerankten Fächern und Hochschulen Ausschau halten. Schon aus Kostengründen wählen viele Studienanfänger(innen) in Deutschland eine Hochschule in der Nähe ihres heimatlichen Wohnorts. Daher werden selbst Rankings mit für die Studienwahl relevanten Informationen nur die ambitioniertesten Studierwilligen ansprechen. Das Phänomen von Rankings macht aber alle Hochschulen selektiver. Nie gab es in Deutschland so viele Fächer mit Numerus Clausus, und die Tendenz ist weiterhin steigend. Dies steht im Widerspruch zu einer Politik der Öffnung der Hochschulen und der Steigerung von Studierendenzahlen.
- (4) Es stellt sich die Frage, ob Rankings den Hochschulen nutzen. Hoch positionierte Hochschulen werden ihr Möglichstes tun, um die Position zu halten; niedriger positionierte Hochschulen

len werden alles daran setzen aufzusteigen. Die Rationalität eines solchen Organisationsverhaltens ist nur eine scheinbare, weil sie zu Isomorphismus, zur Imitation der Besten, führt, statt zu einer Differenzierung von Profilen, die benötigt wird, um der Heterogenität der Studierenden gerecht zu werden.

- (5) Zu fragen ist außerdem, inwiefern Rankings der Politik dienen. In diesem Kontext ist zu beobachten, dass mittlerweile jeder Staat eine oder sogar mehrere seiner Hochschulen unter den Top-50 oder Top-100 in internationalen Rankings sehen will. Es wird dadurch eng an der Spitze. Aber dieses Ziel zeigt, dass Spitzenuniversitäten wiederum zum Indikator für innovative Stärke im ökonomischen Wettbewerb der Nationen werden. Dies ist auch im Rahmen der deutschen Exzellenzinitiative zum Ausdruck gekommen. Dabei ist die Betonung der Korrelation von hoch positionierten Elitehochschulen mit globaler Wettbewerbsfähigkeit allzu vereinfachend und greift zu kurz. Schädlich wird es für ein nationales Hochschulsystem, wenn Ressourcenallokationen mit Rankingplätzen verbunden werden.
- (6) Bezogen auf die Wirkungen von Rankings sind auf der Ebene des nationalen Hochschulsystems drei Aspekte zu akzentuieren: Erstens können Rankings als Symbol nationaler ökonomischer Performanz ein „Wettrüsten“ unter den Hochschulen auslösen, das bis zum Bankrott einiger Hochschulen führen kann. Zweitens begünstigen Rankings die vertikale Stratifizierung, die dann wiederum politische Ressourcenentscheidungen beeinflussen kann, die zu Ungunsten der schwächeren Hochschulen ausfallen. Dies kann die Performanz des Gesamtsystems gefährden. Und drittens muss Hochschulexpansion mit Differenzierung einhergehen. Eine vertikale Stratifizierung ist dabei nur eine mögliche Form. Zu denken ist auch an intra- und inter-institutionelle Differenzierungen sowie Differenzierungen zwischen dem öffentlichen und privaten Hochschulsektor. Die vertikale Stratifizierung produziert wenige Gewinner und viele Verlierer und bewirkt außerdem isomorphe Effekte.
- (7) Für die Ebene der Hochschulen sind vier Wirkungen festzustellen: Erstens beeinflussen Ranking-Ergebnisse die Verteilung von Aufmerksamkeit und Geld. Kritisch zu betrachten ist hier jedoch die Ebene der Analyse, auf der die vermeintlich Besten identifiziert werden, also je nach Ranking die Hochschule als ganze, das Fach bzw. eine Fächergruppe/Fakultät oder aber einzelne Forscher(innen) bzw. Forschergruppen. Möglicherweise sind die Leistungsdifferenzen zwischen einzelnen Wissenschaftler(inne)n innerhalb eines Fachbereichs größer als die zwischen Hochschulen, werden aber durch die Anlage des Rankings nicht mehr als solche sichtbar. Zweitens wissen die Mitglieder der Hochschulen, dass sie im Rahmen der Messungen erfolgreich sein müssen. Folglich versuchen sie, die relevanten Zahlenwerte zu beeinflussen, indem sie beispielsweise einzelne herausragende Forscherinnen und Forscher berufen. Entsprechend hat auch William Locke festgestellt, dass Hochschulleitungen die Logik von Rankings internalisieren und ihre institutionellen Daten so modifizieren, dass sie der Logik der Rankings entsprechen. Im Endeffekt dienen damit aber die Hochschulen den Zwecken der Rankings und nicht umgekehrt, weil Rankings kodifizieren, welche Arten von organisationalem Verhalten legitim sind. Drittens kommt es zu einem Matthäus-Effekt, bei dem Rankings als entscheidende Signale fungieren: Hoch positionierte Hochschulen bekommen mehr Geld, mehr strategische Partner, bessere Wissenschaftler(innen) und Studierende, mehr Reputation etc. Und viertens zeigen sich isomorphe Effekte auch auf der Ebene der Hochschule, weil Rankings mit einer Defizit-Logik arbeiten. Alle Hochschulen unterhalb der höchsten Position versuchen aufzuholen, und zwar allein fokussiert auf die Daten, die für das Ranking relevant sind, unabhängig davon, inwiefern die Professorinnen und Professoren sie für tatsächlich qualitätsrelevant halten.

- (8) Die vorangegangenen Befunde legen die Frage nahe, wie mit Rankings umgegangen werden sollte. Hier besteht das Desiderat, dass die Internationalisierung und Institutionalisierung von Ranking-Logiken besser erforscht werden sollte. Darüber hinaus sind verschiedene Wege denkbar. Einer, der in der jüngsten Vergangenheit häufiger beschritten wurde, ist der, dass einzelne Hochschulen und Fachgemeinschaften die Beteiligung an Rankings ablehnen. Ein weiteres Vorgehen kann darin bestehen, deutlich mehr Rankings zu unterschiedlichen Aspekten zu initiieren, denn diese würden erstens das einzelne Ranking entmachten und könnten zweitens in der Summe unterschiedlichen Hochschulprofilen besser gerecht werden. Nicht zuletzt bedarf es eines neuen Gesellschaftsvertrags über die Rolle von Hochschulen in den Wissensgesellschaften des 21. Jahrhunderts, der weniger auf instrumentellen und stärker auf kognitiv-normativen Formen des Vertrauens beruht.

Bezogen auf AHELO ist darüber hinaus Folgendes anzumerken: AHELO ist bislang eine Machbarkeitsstudie zur Messung von „learning outcomes“. Die Messwerte sollen über Ländergrenzen hinweg verglichen werden. Bei einem solchen Versuch können die sozialen und biographischen Bedingungen der Studierenden nicht in den Blick genommen werden. Problematisch erscheint darüber hinaus, dass Vergleichbarkeit (über artifizielle statistische Aggregate) hergestellt werden soll, die dann faktisch wieder zu Rankings führt. Und Rankings haben die Festlegung von Standards und Aussagen über die Qualität des Lehrens und Lernens zur Folge. Auf diese Weise findet eine weitere Externalisierung der Bestimmung von akademischer Qualität statt, die letztlich doch wieder nur in Minimalstandards ausgedrückt wird. Individuelle Bildungsbiographien kommen dabei deutlich zu kurz.

Literatur

- Hazelkorn, E. (2011).** Rankings and the Reshaping of Higher Education: The Battle for World-Class Excellence. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Locke, W. (2011).** The institutionalization of Rankings: Managing Status Anxiety in an Increasingly Marketized Environment. In: Shin, Jung Cheol / Toutkoushian, Robert K. (eds.): University Rankings: Theoretical Basis, Methodology and Impacts on Global Higher Education. The Changing Academy – The Changing Academic Profession in International Comparative Perspective (3). Dordrecht, The Netherlands: Springer, pp. 201–228.

AHELO als Ranking? Ergebnisse der Diskussion

Susanne In der Smitten

Die Diskussion nahm im Wesentlichen vier Themenbereiche in den Blick: Erstens die Chancen und Grenzen von Rankings im Allgemeinen, zweitens die Frage, inwiefern AHELO eine Nähe zu Rankings aufweist oder aufweisen kann, drittens die Schwierigkeiten, die mit einer Ranking-Perspektive auf AHELO verbunden sind, und viertens erforderliche Voraussetzungen und mögliche Wege, durch die AHELO sinnvoll in einen Ranking-Kontext integriert werden könnte.

- (1) Zunächst bestätigten vor allem die anwesenden Vertreter des CHE-Rankings einen Großteil der Diagnosen des Eingangsstatements zu den Grenzen und Schwächen vorhandener Rankings. Diese trafen vor allem für traditionelle Rankings wie das „Shanghai Ranking“ oder das „Times Higher Education Ranking“ zu. Dabei sei auch zu bedenken, zu welchem Zweck die Rankings begonnen worden seien: Beim „Shanghai Ranking“ etwa habe die Frage im Vordergrund gestanden, wie chinesische Universitäten in der Forschung in naturwissenschaftlichen Fächern im Vergleich zu Hochschulen aus dem Ausland einzuschätzen seien. Entsprechend seien die Indikatoren dann auch vorrangig auf die Erfassung exzellenter Forschung in naturwissenschaftlichen Fächern ausgelegt.

Die Konzeption neuerer Rankings weiche bewusst davon ab. So berücksichtige beispielsweise das CHE-Ranking ausdrücklich nicht nur den Bereich der Forschung, sondern unterschiedliche Leistungsfelder von Hochschulen, um die Vielfältigkeit verschiedener Exzellenzformen in einer horizontalen Differenzierung der Hochschulen verdeutlichen zu können. Zudem würden in einem partizipativen Ansatz die Urteile unterschiedlicher Bezugsgruppen erfragt. Die erfahrungsbasierten Urteile der Akteure würden sodann Faktenindikatoren gegenübergestellt. Das Ergebnis eines solchen mehrdimensionalen Rankings sei keine klare Rangliste, sondern eine wesentlich gröbere Unterteilung in eine Spitzen-, eine Mittel- und eine Schlussgruppe, die durchaus geeignet sei, auch Studieninteressierten Informationen zu ihrer Orientierung an die Hand zu geben.

In der Vergangenheit habe die Nutzung von Rankings manche fragwürdige Facette gezeigt, etwa durch Maßgaben an Hochschulen, nur noch mit Top-Universitäten zu kooperieren, durch die Beschränkung von Stipendien auf Studierende hoch geranker Hochschulen oder durch Verbindungen von Rankingergebnissen zu Mittelvergabesystemen. Zwar sei es zu begrüßen, dass politische und strategische Entscheidungen stärker evidenzbasiert getroffen würden. Daten aus Rankings nähmen solche Entscheidungen aber nicht vorweg, sondern bedürften in diesem Prozess stets der Reflexion und Interpretation und müssten zu politischen Zielen ins Verhältnis gesetzt werden.

- (2) Die Diskussion verdeutlichte, dass die Verbindung von AHELO zu Rankings insofern naheliegend sei, als im Rahmen der AHELO-Machbarkeitsstudie systematisch statistische Daten über den Kompetenzerwerb Studierender generiert wurden. Die Zahlenwerte aus der Studie auszuweisen, diene der Vergleichbarkeit, führe aber automatisch auch zu der Möglichkeit zur Bildung einer Rangfolge und damit zu einem Ranking. Eine solche Reihung habe den Vorteil, durch eine Reduktion von Komplexität eine rasche Orientierung zu erleichtern. Dadurch könne im positiven Sinne auch ein vorhandener Handlungsbedarf zur Verbesserung leichter erkannt werden. Die allen vergleichenden Untersuchungen dieser Art inhärente Gefahr einer

einseitigen und verzerrenden Ergebnisdarstellung im Kontext von Rankings sei zunächst generalisierend zu konstatieren, jedoch im Zusammenhang mit den möglichen Zielen von Adressaten einer AHELO-Berichterstattung konkret zu diskutieren.

Hervorzuheben sei, dass bisherige Rankings vor allem Rahmenbedingungen der Leistungen von Hochschulen erfassten, während AHELO stärker auf das inhaltliche Geschehen in der Lehre ausgerichtet sei und dadurch ein wertvolles neues Element bedeuten könne.

- (3) Allerdings bilde auch AHELO, sofern es gelinge, einen Kompetenzfortschritt der Studierenden im Verlauf ihres Studiums sichtbar zu machen, wiederum nur eine bestimmte Facette der Leistungen von Hochschulen ab.

Zudem handele es sich bei der Kompetenzorientierung um einen Bereich der Bologna-Reformen, der bislang noch wenig verstanden sei. Zwar seien Hochschulangehörige, nicht zuletzt durch die Akkreditierungsverfahren, inzwischen darin geübt, Kompetenzziele und „learning outcomes“ bezogen auf Studiengänge, Module und Veranstaltungen zu formulieren. Das Wissen darüber, wie Lehre und Prüfungen kompetenzorientiert ausgerichtet werden könnten, wie also die gewünschten Outcomes generiert und bewertet werden könnten, sei demgegenüber noch deutlich weniger entwickelt. Hier bedürfte es eigentlich breit angelegter hochschuldidaktischer Weiterbildungen. Im Ergebnis messe AHELO Aspekte, auf die die Hochschullehre möglicherweise bislang noch nicht hinreichend ausgerichtet sei. Ein konzeptionelles Manko sei zudem die Art, wie die Messung erfolge, insbesondere hinsichtlich der Trennung von generischen und fachlichen Kompetenzen. Da auch generische Kompetenzen in Verbindung mit fachlichen Inhalten erworben und erprobt würden, sei eine stärker fachgebundene Erhebung sinnvoll.

Eine zu starke Fokussierung auf AHELO berge die Gefahr, die Relevanz von Kompetenzen als „learning outcomes“ bei der Beurteilung der Qualität von Hochschulen zu überhöhen. Dies gelte umso mehr, als der Qualitätsbegriff im Hochschulbereich nicht klar formuliert sei und die OECD als Initiator von AHELO den Charakter eines Akteurs mit hoher Autorität habe, dessen gleichsam amtliche Untersuchungsergebnisse einen hohen Druck zur Standardisierung erzeugten. Die Frage, inwiefern dieselben Kompetenzen in allen Ländern und an allen Hochschulen gleichermaßen relevant seien, trete dann möglicherweise in den Hintergrund. Wenn die Lehre an den verschiedenen Standorten sich aber zunehmend angleiche, könne dies auch den Anreiz für Studierende zur Mobilität reduzieren.

- (4) Zuletzt wurde erörtert, was geschehen müsse, um AHELO in einem Ranking-Kontext sinnvoll nutzen zu können. Hier wurde hervorgehoben, dass es zunächst wichtig sei, die Zielsetzungen der Untersuchung klarer zu formulieren und die Instrumente zu optimieren. Insbesondere sollten generische und fachliche Kompetenzen auf der Fächerebene verbunden werden. Darüber hinaus sollten die Kompetenzmessungen nicht für sich alleine stehen, sondern entsprechend den Ergebnissen des internationalen U-Multirank-Projekts in einen multidimensionalen Kontext eingebettet werden. Die im Rahmen eines Studiums erweiterten Kompetenzen könnten etwa in der Zusammenschau mit Zufriedenheitswerten der Studierenden und quantitativen Indikatoren zu Fakten wie Studienabbruchquoten durchaus Einblicke in die Qualität der Lehre vermitteln. Dabei sollten die Ursachen der Befunde und Bezüge zwischen den verschiedenen Dimensionen nach Möglichkeit erforscht und verdeutlicht werden.

Ein höchst relevanter Punkt sei zudem die Kommunikation der Ergebnisse: Hier sollten die Zielgruppen klarer definiert und die Darstellung der Ergebnisse auf sie ausgerichtet wer-

den. Um eine Simplifizierung zu vermeiden, sei ein Nutzerguide für verschiedene Gruppen, etwa für Studierende oder politische Akteure, wünschenswert. Eine Verbindung der AHELO-Ergebnisse zur Mittelvergabe an die Hochschulen sollte generell unterbleiben.

„AHELO goes Germany“ – aber nur, wenn Ziele geklärt, Methoden optimiert und Risiken eingedämmt sind

Frank Ziegele

Zu den Zielen von AHELO

Eine positive Entscheidung Deutschlands, sich an AHELO zu beteiligen, sollte erst getroffen werden, wenn die Zielsetzung und die Zielgruppe von AHELO klar formuliert ist.

Problematisch ist, dass zum jetzigen Projektstand offenbar noch immer einige elementare Parameter des Konzepts unklar sind. So ist eine der wichtigsten Setzungen noch offen: Was ist das eigentliche Ziel, das mit AHELO verfolgt werden soll? Welche Funktion soll es übernehmen – die eines Lerninstruments für Hochschulen, eines Benchmarking-Ansatzes für Hochschulen, die sich vergleichen wollen, oder doch gar die eines Rankings im internationalen Vergleich? Entsprechend ist auch die Zielgruppe unklar: Soll es ein Instrument speziell für die einzelnen Hochschulen sein? Oder ein Transparenz-Ansatz für eine breite Öffentlichkeit? Hier bedarf es erst eindeutiger und verbindlicher Klärungen, bevor ein Commitment zur Teilnahme Deutschlands an AHELO zu befürworten ist. Denn nicht alle Zielsetzungen erscheinen gleichermaßen sinnvoll, wie die folgenden Überlegungen zeigen.

Sinnvoll erscheint die Analyse von „learning outcomes“ als Lerninstrument für die einzelne Hochschule in Bezug auf die Gestaltung der Lehre. Kontextualisierung geht dann vor Vergleichbarkeit der Ergebnisse.

Wenn AHELO der einzelnen Hochschule relevante Informationen erbringen soll, sind zwei Herangehensweisen denkbar: Man könnte entweder den „value added“ betrachten oder das Erreichen eines spezifischen Kompetenzsolls. Beides würde empirisch gesicherte Werte erfordern und relevante Schlussfolgerungen aus den Lernergebnissen für den Erfolg und die Gestaltung der Lehre ermöglichen.

Der „value added“, d. h. der Zuwachs an Kompetenzen, ermöglicht durch das Messen der individuellen Eingangskompetenzen der Studierenden im Vergleich mit den Kompetenzen am Ende des Studiums die Kontrolle der Entwicklung. Das wird aber in der AHELO Machbarkeitsstudie, die kein Paneldesign aufweist, gar nicht gemacht – daher bleibt die Aussagefähigkeit der Pilotstudie für diese wichtige Frage begrenzt.

Es kann aber auch anhand eines festgelegten Soll-Zustands der Lernerfolg an einer Hochschule bewertet werden. Hierbei würde für jedes Fach an der spezifischen Hochschule ein zu erreichendes Kompetenzniveau als Referenzmaßstab festgelegt und mit Hilfe eines Soll-Ist-Vergleichs überprüft, ob dieses Ziel erreicht wurde. Allerdings erscheint die spezifische Operationalisierung solcher Kompetenz-Zielgrößen für einzelne Hochschulen, auf die Testergebnisse dann bezogen werden können, nicht trivial. Sie würde ein hochgradig flexibles, modularisiertes Testsystem erfordern, das sich den Bedürfnissen der Hochschulen anpassen lässt. Auch dies ist im derzeitigen AHELO-Projekt nicht erprobt worden.

Eine solche Kontextualisierung erscheint dennoch als wichtiger Schritt und ermöglicht, dass die Kompetenzmessung für die Optimierung der Lehre an der jeweiligen Hochschule einsetzbar ist. Dann wird aber die internationale Vergleichbarkeit deutlich eingeschränkt, da sich Ergebnisse immer nur auf das spezifische, zu erreichende Kompetenzniveau beziehen. D. h. nicht alle mögli-

chen Ziele sind gleichzeitig erreichbar, der direkte internationale Vergleich wäre nach dieser Vorstellung nachrangig. AHELO würde zu einem Serviceangebot von Werkzeugen, die Hochschulen nach ihren Interessen einsetzen können. Damit würde sich AHELO fundamental von den Schulvergleichen à la PISA unterscheiden.

Ein weltweiter Vergleich ist als passendes Ziel für AHELO fraglich, da kulturelle und fächer-spezifische Unterschiede ausgeblendet werden.

Es ist nach wie vor fraglich, wie sinnvoll es wirklich ist, „learning outcomes“ international zu vergleichen. Kann dies ohne kulturellen Bias gelingen und wenn ja, unter welchen Bedingungen? Im Unterschied zum Schulbereich, in dem – zumindest bei den PISA-Messungen – eher grundlegende Kulturtechniken (Lesen, Rechnen, Textverständnis) geprüft werden, die mehr oder weniger universell sind, haben wir es im Hochschulbereich zumindest in einem Teil der Fächer mit sehr unterschiedlichen nationalen Wissenschaftstraditionen zu tun – auch im Sinne von Profilen/ Profilbildung. Gerade deshalb erscheint der kontextualisierte, hochschulspezifische Ansatz plausibler als der international komparative. Äußerungen vieler Teilnehmer an AHELO gehen dennoch weiter in Richtung des internationalen Vergleichs.

Die spezifische Vielfalt innerhalb von und zwischen Hochschulsystemen wird durch eine Festlegung von fachlichen Kompetenzen im Sinne eines Kanons, der gemessen werden soll (und das muss zwangsläufig eine begrenzte Auswahl sein), eingeebnet. Es erscheint zweifelhaft, ob im tertiären Bildungsbereich in allen Fächern ein Konsens über einen länder- und kulturübergreifenden Kanon an Kompetenzen hergestellt werden kann. Dies gilt in besonderem Maße für geistes- und sozialwissenschaftliche Fächer, die stark an die jeweiligen Denktraditionen gebunden sind, aber auch für technisch-naturwissenschaftliche Fächer, wie Rückmeldungen von international rekrutierenden Arbeitgebenden zeigen, die für spezifische Kompetenzprofile gezielt auf solche Differenzierungen zurückgreifen. Wenn ein Arbeitgeber als Ingenieur gezielt einen Tüftler sucht, einen Erfinder oder aber einen Manager, wird er sich gezielt Absolvent(inn)en aus bestimmten Hochschulsystemen/Studiengängen aussuchen, die stärker auf eines der Profile ausgerichtet sind. Diese Vielfalt sollte erhalten bleiben und berücksichtigt werden. Weltweite Vergleiche sind damit nicht ausgeschlossen, müssen aber so gestaltet sein, dass sie Kompetenzvielfalt fördern. Auch hierzu liegen bei AHELO noch keine hinreichenden Ergebnisse vor.

Zur Methodik

Erhebungsinstrumente, die in einem bestimmten Kontext entwickelt wurden, können nicht ohne kritische Überprüfung/Überarbeitung auf andere Konstellationen übertragen werden. Bisher ist nicht deutlich geworden, dass dies in der feasibility study hinreichend berücksichtigt worden ist.

Es ist zu prüfen, inwieweit Kompetenzen (die nicht unabhängig von Lehr- und Lernkulturen sind) weltweit mit einheitlichen Instrumenten gemessen werden können, die aus einem bestimmten institutionellen oder nationalen Kontext kommen. So wissen wir noch zu wenig über internationale Unterschiede zwischen nationalen Prüfungs-Traditionen, z. B. bezüglich schriftlicher vs. mündlicher „Präsentation“ von Kompetenzen durch Studierende und die zugrunde liegenden Lehr- und Lernkulturen; z. B. hinsichtlich der Bedeutung von und Vertrautheit mit Einzelarbeit vs. Teamwork, schriftlicher Darstellung vs. Multiple Choice etc. Erhebungsinstrumente, die auf die Ausrichtung einer Hochschulform in einem bestimmten Hochschulsystem passen, müssen deshalb bei der Übertragung in andere Hochschulsysteme/auf andere Hochschulformen kritisch ge-

prüft und gegebenenfalls auch mit Blick auf die „Arten“ der Messung/Prüfung angepasst werden. Wenn z. B. zur Messung der „generic skills“ bei AHELO ein Verfahren verwendet wird, das im Kontext der US-amerikanischen Liberal Arts Colleges entwickelt wurde, erscheint dessen genereller Einsatz für alle Fächer (in der feasibility study z. B. für Ingenieure) und für alle Kulturen und Hochschultypen fraglich. Auch bei den „generic skills“ gibt es eine Bandbreite, die in stärker fachlich geprägten Studienprogrammen fächerspezifisch variieren sollte.

Die reine Betrachtung von „learning outcomes“ greift zu kurz – objektive Daten zum und individuelle Zufriedenheit mit dem Studium müssen auch betrachtet werden.

Die Messung von „learning outcomes“ und eventuell sogar des „value added“ ist so etwas wie der „heilige Gral“ der Qualitätsbewertung – häufig wird argumentiert, dies sei das eigentlich Relevante und alle anderen Messansätze und Indikatoren in der Lehre seien nur weniger geeignete „proxies“.

Wenn aber Hochschulen Lehre und Studium verbessern wollen, brauchen sie noch mehr Informationen. Diese sollten den gesamten Ablauf des Lehrens und Lernens abbilden, von der Qualität der Inputs über die Prozesse bis zu den Ergebnissen. Nur ein multidimensionaler Ansatz der Messung kann der Vielfalt der Einflussfaktoren auf gute Lehre gerecht werden. Ein hoher Lernzuwachs ist z. B. in den USA für Community Colleges mit bildungsfernen Zielgruppen wichtig, die Studierende mit niedrigem Kompetenzerwerb in der sekundären Bildung aufnehmen und diese zu einem höheren Niveau führen wollen. Für eine „MBA-Schmiede“ mit handselektierten, bereits zu Beginn hochqualifizierten Studierenden ist das Ausmaß des Lernzuwachses möglicherweise nicht das zentrale Thema, sondern der Arbeitsmarkterfolg. Wenn Hochschulen „learning outcomes“ mit einer Ursachenanalyse in Verbindung bringen wollen, dann brauchen sie Studierendenurteile über Betreuungssituation, Studienangebot u.ä. Insgesamt sind das keine Argumente gegen AHELO, sondern Argumente dafür, dass Kompetenzzuwachs nur eine Zielgröße unter mehreren in einer multidimensionalen Darstellung von Hochschul-Performance sein sollte, die für unterschiedliche Profile von Hochschulen unterschiedliche Relevanz hat, aber nicht zum Königsweg der Messung von Lehr-Performance deklariert werden sollte. Weitere relevante Dimensionen sind objektive Daten, wie etwa Fakten zu Studienabbruchquoten, Studiendauer, aber auch die individuelle Zufriedenheit der Studierenden mit ihrem Studium bzw. ihrer „learning experience“. Es wäre also falsch, „learning outcomes“ als einzig relevante Größe der Qualität von Lehre und Studium zu verabsolutieren; vielmehr sollten sie den anderen genannten Messansätzen im Rahmen von Analysen gegenübergestellt werden.

Wenn man also tatsächlich AHELO auch für internationale Vergleiche nutzen will, resultieren nun zusammenfassend aus den bisher diskutierten Punkten folgende Anforderungen: (1) Das System findet geeignete Referenzgrößen (value added, Kompetenz-Soll), (2) es ermöglicht je nach Fach vielfältige Kompetenzprofile; (3) Messansätze sind an kulturelle und fachliche Bedingungen flexibel anpassungsfähig; (4) das System ordnet die „learning outcomes“ in einen multidimensionalen Messkontext ein. Ob das alles zusammen funktioniert, scheint derzeit noch nicht geklärt.

Zur Operationalisierung

Die Vorträge haben einige bedenkliche Methodenfragen offenbart.

Die Vorträge im Rahmen des Workshops haben zahlreiche methodische Fragen berührt. Es soll nur exemplarisch auf einige verwiesen werden, die als problematisch zu werten sind, und gezeigt werden, dass bei der Ausarbeitung der Methoden noch ein langer Weg vor uns liegt:

Vergleichende Transparenz- und Messinstrumente müssen unbedingt in all ihren Methoden selbst transparent sein. Bei AHELO erscheint das z. T. nicht gegeben zu sein, z. B. in Bezug auf „confidential items“. Auch wenn AHELO kein Ranking sein will, sollten doch die diesbezüglichen Standards der IREG-„Berlin Principles“⁷ beachtet werden, die für Transparenzinstrumente im weiteren Sinn sinnvoll sind.

In der Machbarkeitsstudie lag die Teilnahmequote der Studierenden in manchen Ländern bei nur 10% - aber es geht nicht nur um ein „mehr“ an Studierenden, sondern auch um einen repräsentative Auswahl, die durch Selbstselektionseffekte möglicherweise nicht gewährleistet ist.

Kulturelle Spezifika können Testergebnisse beeinflussen. In der jetzigen Operationalisierung von AHELO können Artefakte entstehen, wie ein kleinteiliges Beispiel zeigt: So beantworten die Studierenden die Fragen am Computer und haben daher weltweit unterschiedliche Voraussetzungen, das Tippen ist auf arabisch eingestellten Tastaturen zeitaufwändiger als etwa bei deutsch- und englischsprachigen Tastaturbelegungen. Daher müsste Studierenden ein unterschiedlich langer Zeitraum für das Ausfüllen zur Verfügung gestellt werden. Dies ist nur ein möglicher Faktor, der zu Verzerrungen führen könnte – illustriert aber die Schwierigkeiten, vor denen die internationale Anwendung von AHELO steht. Wobei sich diese Probleme zum Teil relativieren, wenn beim Anspruch auf internationale Vergleichbarkeit Abstriche gemacht werden (siehe die obige Diskussion über die Ziele).

Zu möglichen Risiken

Durch das Engagement der OECD könnte dem Vorhaben ein offizieller Anspruch zugesprochen werden.

Der Impact durch die OECD als Verantwortliche für AHELO sollte beobachtet werden: Hierdurch wird AHELO als das offizielle Messsystem der OECD und als „Hochschul-PISA“ mit entsprechendem Impact betrachtet werden. Den Ergebnissen würde daher ein Wahrheitsanspruch angetragen werden. Eine vergleichbare Konstellation gab es bei U-Multirank: Hier möchte sich die Europäische Kommission gerade wegen dieser Problematik mittelfristig aus dem Projekt herausnehmen, also kein „offizielles“ EU-Ranking entwickeln, sondern lediglich eine Anschubfinanzierung zur Entwicklung eines unabhängigen Systems leisten. Da die OECD hier wohl kein Problem sieht, sucht sie auch nach keiner Lösung.

Die ausführliche Datenbasis von AHELO könnte als Möglichkeit eines eindimensionalen Rankings missverstanden werden.

Es bleibt fraglich, ob die Beteuerung der OECD, dass es sich um kein Ranking handle und keine Betrachtungen auf nationaler Ebene erfolgen sollen, tatsächlich durchsetzbar ist. Liest man beispielsweise die Begründung einer australischen Vertreterin auf der AHELO-Website, warum sich Australien beteiligt, dann steht dort „the OECD study could provide a useful complement to other higher education ranking systems that are based mainly on research and/or reputation“⁸. Das Vorhandensein einer Datenbasis wie AHELO wird – im Erfolgsfall – zwangsläufig „Ranking-Begehrlichkeiten“ wecken; dem wird sich die OECD nur entziehen können, wenn die Ergebnisse nicht als Hochschulvergleich veröffentlicht werden. Damit wird aber wiederum der öffentliche Nutzen von AHELO als Transparenzinstrument deutlich eingeschränkt, was auch mit Blick auf die außerordentlich hohen Kosten problematisiert werden kann.

⁷ http://www.ireg-observatory.org/index.php?option=com_content&task=view&id=107&Itemid=99

⁸ <http://www.oecd.org/edu/highereducationandadultlearning/ahelo-theviewfromaustralia.htm>

AHELO im Spannungsfeld von
Wissenschaft, Gesellschaft und Bildungspolitik

AHELO: Paradigmenwechsel oder mehr vom Gleichen?

Stefan Hornbostel

Einleitung

An Kompetenzmessungen, die neben manch analytischen Einsichten am Ende internationale oder auch nationale Ranglisten produzieren, hat man sich inzwischen gewöhnt. Zumindest Schülern und Eltern ist die PISA-Studie vertraut, die seit dem Jahr 2000 von der OECD durchgeführt wird. PISA misst die Kompetenzen von Schülern und kann dabei sowohl auf (umstrittene, aber tendenziell konsensfähige) Standards als auch – dank der Kooperation staatlicher Schulbehörden – eine repräsentative Sampling-Strategie verfolgen (auch wenn die Stichprobenziehung äußerst kompliziert und wenig transparent ist). Viele weitere Tests mit klangvollen Namen wie TIMSS, PIRLS, ICCS, IGLU, DESI oder VERA widmen sich speziellen Kompetenzbereichen von Schülern und Lehrern.

2008 begannen die Planungen für ein weiteres OECD-Projekt: PIAAC (Programme for the International Assessment of Adult Competencies) soll sich basalen Kompetenzdimensionen, die zur Bewältigung alltäglicher und beruflicher Anforderungen von Erwachsenen notwendig sind, mithilfe üblicher Bevölkerungsstichproben widmen. Erste Ergebnisse werden zum Ende dieses Jahres erwartet.

Nach einer ersten Absprache der Bildungsminister 2006 in Athen wurde schließlich ebenfalls im Jahr 2008 in Tokyo ein weiteres Kompetenzmessungsprojekt beschlossen: das Assessment of Higher Education Learning Outcomes (kurz: AHELO). Dabei handelt es sich zunächst nicht um eine Vollerhebung, sondern um eine Machbarkeitsstudie, die Aufschluss darüber geben soll, ob international vergleichend die Kompetenzen von Studierenden nach drei bis vier Jahren Studium so erfasst werden können, dass einerseits Aussagen über Struktur und Ausmaß der erworbenen Kompetenzen gemacht werden können, andererseits internationale Vergleiche ohne kulturellen oder sprachlichen Bias möglich werden. Das Projekt war mit kürzeren Entwicklungszeiten und unsicherer Finanzierung noch ambitionierter als PIAAC angelegt, was dem Vorhaben nicht gut getan hat.

Dass dieses Vorhaben nicht nur auf Interesse, sondern auch auf Skepsis stieß, zeigt sich bereits an der Liste der teilnehmenden Staaten: 17 Länder nehmen teil, von denen viele erst im Laufe des Projekts dazu stießen (also an der Instrumentenentwicklung kaum teilhatten), ist zwar zunächst eine beeindruckende Zahl. Die Liste der Teilnehmer (Abu Dhabi, Australien, Belgien, Kanada (Ontario), Kolumbien, Ägypten, Finnland, Italien, Japan, Korea, Kuwait, Mexiko, Niederlande, Norwegen, Russische Föderation, Slowakei sowie einige US-amerikanische Universitäten) ist aber mindestens so interessant wie die Liste der wissenschaftlich bedeutsamen Länder, die nicht teilgenommen haben, wozu auch die Bundesrepublik Deutschland gehört. Dabei waren weniger die Kosten von rund neun Mio. Euro (ohne die nationalen Aufwendungen) ausschlaggebend, als vielmehr grundsätzliche Fragen und Zweifel, wie sie bereits um die PISA-Studie aufgekommen waren.

Einige dieser Fragen sollen im Folgenden angesprochen werden. Da Geschichte, Design, Instrumentierung und erste Ergebnisse der AHELO-Studie inzwischen in zwei Berichtsbänden (OECD, 2012; OECD, 2013) dokumentiert sind und ein dritter Band mit weiteren Ergebnissen für dieses Jahr angekündigt ist, kann hier auf eine Darstellung des AHELO-Projekts verzichtet werden. Stattdessen sollen einige Fragen, die auch in der das Projekt begleitenden Gruppe nationaler Experten (GNE) diskutiert wurden, angesprochen werden – dies vor dem Hintergrund, dass die Machbar-

keitsstudie Ende 2012 beendet wurde und nun über eine – wie auch immer geartete – Fortsetzung als Vollerhebung oder weitere Vorlaufstudie zu entscheiden sein wird.

Wer sind die Adressaten von AHELO?

Soviel ist zunächst sicher: Das Anliegen von AHELO ist erstens anders als in den übrigen OECD-Assessments, zweitens unklarer und drittens weitaus schwieriger zu realisieren. Unmittelbar deutlich wird das in vielen Begleittexten zu AHELO, die zunächst einmal klären, was AHELO nicht sein will. Die Verwirrungen resultieren vor allen Dingen aus der sehr breit definierten Zielgruppe: Arbeitgeber sollen über die Fähigkeiten von Absolventen informiert werden, Regierungen darüber, ob ihre Investitionen in die tertiäre Bildung gut und effizient sind, Hochschulen und Fakultäten sollen eine Datengrundlage für die Verbesserung ihres Ausbildungsangebots und Studierende eine bessere Entscheidungsgrundlage für die Wahl einer Hochschule erhalten (OECD 2011: 10). Je nachdem welche dieser Ziele man in den Vordergrund rückt, entstehen unterschiedliche Designanforderungen an das Assessment:

- Die Unterstützung bei der Studienwahl verlangt eine möglichst flächendeckende Erhebung, die in ein Rating oder Ranking der Institutionen mündet – etwas, was AHELO explizit nicht produzieren will. Individuelles Feedback an Studierende derart, dass Studierende das Assessment als individuelles Diagnoseinstrument nutzen können, um die Ergebnisse in ihren Entscheidungsprozessen zu berücksichtigen, ist in AHELO nicht vorgesehen.
- Die Information von Regierungen setzt voraus, dass die Ziele der Bildungsinvestitionen geklärt sind: Geht es um die Menge des im internationalen Vergleich hochqualifizierten Nachwuchses, geht es um einen möglichst großen Kompetenzzuwachs der Studierenden gemessen am Eingangsniveau, geht es um die Förderung von Problemgruppen, Senkung von Abbruchquoten oder spezifische Arbeitsmarktqualifikationen? Ungelöst ist auch die schon in PISA aufgeworfene Frage, unter welchen Bedingungen die große und tendenziell größer werdende Binnenvarianz in den Teilnehmerstaaten im Hinblick auf Ziele und Profile der tertiären Bildungseinrichtungen einen sinnvollen internationalen Vergleich erlaubt.
- Die Information der Institutionen setzt voraus, dass eine geeignete Referenzpopulation definiert werden kann. Institutionelle Klassifikationen, die über die Carnegie Classification hinausgehen wie zum Beispiel das europäische U-Multiranking stecken aber noch in den Kinderschuhen. Insofern ist es schwer, geeignete Subgruppen zu definieren. In AHELO wurde dieses Problem durchaus gesehen und mit einer eigenen Erhebung der institutionellen Kontextbedingungen bearbeitet. Ein Ersatz für eine elaborierte Klassifikation ist dies allerdings nicht. Der institutionelle Fokus wurde im Laufe des AHELO-Projekts immer dominanter und ist sicherlich auch die vielversprechendste Perspektive, stellt allerdings erhebliche Anforderungen an die Sampling-Strategie, da in diesem Fall die Studienabschlusskohorte innerhalb der Hochschule die jeweilige Grundgesamtheit darstellt.
- Arbeitgeber schließlich führen für hochqualifizierte Jobs längst eigene Assessments durch. Hier dürfte das Interesse entweder auf spezifische Fähigkeit gerichtet sein oder auf Informationen über Institutionen, die bisher nicht im Blickfeld waren. Für Rekrutierungsprozesse oder interne Fortbildungsprogramme dürfte AHELO aber nicht allzu viele neue Erkenntnisse bringen.

Eine Weiterentwicklung von AHELO wird also in diesen Punkten größere Klarheit brauchen.

Was ist das (Erkenntnis-)Ziel?

Die Grundüberlegungen, die hinter AHELO standen, sind ebenfalls vielfältig:

- Grundsätzlich verändern weltweit die steigenden Partizipationsraten an der tertiären Bildung die Funktion der Hochschulen. Ein immer kleinerer Teil der Absolventen verbleibt im Rahmen von Selbstrekrutierungsprozessen im Wissenschaftssystem, ein immer größerer Teil wird in seiner weiteren beruflichen Tätigkeit nicht unmittelbar wissenschaftlich tätig sein. Das wirft die Frage nach erwarteten und vorhandenen Kompetenzen auf.
- Seit Langem ist erkennbar, dass die im Rahmen des New Public Managements vorgenommene Verschiebung von einer Input- zu einer Output-Steuerung auch die Hochschullehre erfassen wird. Bisher können aber im Wesentlichen nur Hilfskonstruktionen wie Abbruchquoten, Zufriedenheitsmessungen oder Absolventenstudien herangezogen werden. AHELO wäre insofern eine überfällige Ergänzung existierender Rankings, die versuchen, Lernchancen über Rahmenbedingungen zu approximieren und insoweit Teil der accountability.
- Reputationszuweisungen innerhalb des Wissenschaftssystems erfolgen über Forschungsleistungen. AHELO setzt hier einen deutlichen und sicherlich notwendigen Akzent auf die Lehre und die Ausbildungsqualität. Bereits bekannt ist, dass Hochschulen mit starken Forschungsleistungen keineswegs automatisch auch ein entsprechend gutes grundständiges Ausbildungsangebot vorhalten.
- AHELO steht allerdings nicht in einer hochschuldidaktischen Traditionslinie. Das Assessment fragt nach „*learning outcomes*“, nicht nach „*teaching outcomes*“. Auch dies ist ein begrüßenswerter Perspektivenwechsel angesichts der Tatsache, dass die Lernchancen vielfältiger geworden sind und sich nicht zuletzt durch die elektronischen Medien tendenziell von unmittelbarer Präsenz in der Institution ablösen. Das allerdings macht auch die Zurechenbarkeit von Kompetenzen auf Organisationsleistungen schwieriger.
- Diese Perspektive provoziert aber auch die Frage nach dem Ziel des Assessments. Geht es um die Feststellung einer Organisationsleistung im Sinne eines value added Ansatzes oder geht es um die Beschreibung der Verteilung des Kompetenzniveaus in den Abschlusskohorten, unabhängig von den Eingangsqualifikationen der Studierenden. Je nach Zielgruppe sind beide Informationen von Interesse. Ein in AHELO vorgesehene Modul zur Entwicklung eines value added Messmodells konnte wegen der zeitlichen und finanziellen Restriktionen nicht umgesetzt werden, wäre allerdings in Europa auch kaum zu realisieren gewesen, da kaum Verfahren zur Messung der Eingangsqualifikation eingesetzt werden.
- AHELO trägt mit seiner international vergleichenden Perspektive schließlich einer, wenn auch nur langsam wachsenden Mobilität von akademisch ausgebildetem Personal und einer zunehmenden Internationalisierung der Arbeitsmärkte für hochqualifiziertes Personal Rechnung.
- Das wichtigste – und vermutlich auch umstrittenste – Ziel aber ist die Etablierung einer Kompetenzmessung komplementär zu den Leistungszertifizierungen der Hochschulen. Noten und Zertifikate sind meist bereits national nicht vergleichbar, erst recht aber nicht international. Inflationäre Bestnotenvergaben oder lokale Idiosynkrasien bei der Benotung, wie sie in Deutschland zu beobachten sind, (vgl. Wissenschaftsrat, 2012, Hornbostel & Lottmann, 2012) mindern die Aussagekraft derartiger Zertifikate weiter. Die Etablierung eines internationalen Standards kann man sicherlich – wie viele Kritiker vorbringen – als Einschränkung hochschulischer Autonomie sehen. Man kann sie aber auch als Rückgewinnung von Autonomie verste-

hen, da die sinkende Aussagekraft von Zertifikaten notwendig dazu führt, dass die Kompetenzfeststellung aus der Hochschule auswandert.

Es sprechen also viele Gründe dafür, sich dem Thema der Kompetenzmessung in der Hochschulbildung zuzuwenden. Auch die grundsätzliche Anlage als international standardisierte Messung ist überzeugend. Voraussetzung ist allerdings die Entwicklung eines validen und reliablen Instruments.

Kann man Kompetenzen von Studierenden messen?

Die Grundsatzdebatte, ob eine Messung von Kompetenzen überhaupt möglich ist und welche Folgen dies hat, ist bereits in der PISA-Debatte ausführlich geführt worden und muss hier nicht wiederholt werden. Für die tertiäre Bildung kommen aber einige Aspekte hinzu. Der in Europa in allen Staaten zu verzeichnende Trend, die hochschulische Autonomie zu erhöhen, die Einrichtungen zur stärkeren Ausbildung je eigener Profile zu animieren, die Durchlässigkeit des Systems zu erhöhen, neue Berufsfelder zu akademisieren, und schließlich der traditionell hohe Freiheitsgrad, der mit einer Persönlichkeitsbildung durch forschendes Lernen verbunden wird, scheinen zunächst gegen eine einheitliche Kompetenzmessung zu sprechen. Andererseits haben wir eine bis in die 1970er Jahre zurückreichende Tradition in Gestalt von Rahmenprüfungsordnungen, durch Fachgesellschaften definierte curriculare Mindeststandards, in Akkreditierungsverfahren definierten Mindestanforderungen sowie im Rahmen des Tuning-Projekts (www.unideusto.org) auch erste ausformulierte disziplinäre Kompetenzanforderungen auf europäischer Ebene. Insoweit ist eine Kompetenzmessung völlig kompatibel mit den disziplinären Bemühungen, sich auf einen international gültigen Katalog von Kompetenzen eines Hochschulabsolventen zu einigen. Ein Messkonstrukt kann man sich entsprechend wie eine nach oben offene Richterskala vorstellen, die im unteren und mittleren Bereich vermutlich standardisierbar, im Hinblick auf je besondere Fähigkeiten aber kaum aussagekräftig ist, also lediglich einen Teil, aber einen sehr wichtigen Teil der Kompetenzstruktur erfassen kann. Dass keineswegs in allen Disziplinen der Diskussionsprozess über eine derartige „baseline“ ausreichend weit vorangeschritten ist, spricht ebenso wenig gegen die Entwicklung eines Messverfahrens wie die Tatsache, dass zweifellos die lange Tradition von Leistungstests im anglo-amerikanischen Kulturraum gelegentlich imperiale Auswirkungen auf die Instrumentenentwicklung hat. Vielmehr hat AHELO einen überfälligen Diskurs innerhalb der Disziplinen stimuliert, die Formulierung von Kompetenzstrukturen in eigener Regie voranzutreiben. Dies könnte auch den Austausch zwischen den Disziplinen fördern. So liegen in einigen Disziplinen – wie etwa der Medizin – bereits jetzt viele Erfahrungen mit standardisierten Tests zur Lernstandserfassung und zum Kompetenzerwerb vor (Osterberg, Kölbel & Brauns, 2006). Die Erfahrungen mit diesen Instrumenten könnten wertvolle Informationen für andere Disziplinen sein. Gerade hinsichtlich der Frage, in welchem Maße neben fachspezifischen Kompetenz auch „generic skills“ als eigenständige, disziplinunabhängige Kompetenz erfasst werden sollen und können, war im AHELO-Projekt einer der umstrittensten Teile der Projektentwicklung. Unstrittig war dabei, dass grundlegende allgemeine Kompetenzen wie kritisches Denken, Argumentationsfähigkeit, Problemlösungskompetenz, Kommunikationsfähigkeit etc. zur Kompetenzstruktur eines Akademikers gehören, strittig war allerdings, inwieweit sich diese Kompetenzen von den disziplinären Kontexten lösen lassen. Die Meinungsbildung war in diesem Punkt innerhalb der Expertengruppe deutlich europäisch geprägt und folgte weniger dem amerikanischen Modell einer allgemeinen Erziehungsaufgabe der Hochschule als vielmehr einer disziplinären Orientierung. Auch im Hin-

blick auf die technische Umsetzung dieser Messung war AHELO lehrreich. Die Adaption des amerikanischen CLA-Tests (www.collegiatelearningassessment.org) erwies sich nicht nur als aufwendig, sondern auch als anfällig für kulturelle Bias-Faktoren und problematisch sowohl hinsichtlich der Validität wie der Reliabilität. Multiple-Choice-basierte Alternativen scheinen aber durchaus möglich zu sein. Derartige Befunde erwartet man von einer Machbarkeitsstudie, deren Ziel nicht vorrangig die Auslieferung eines fertigen Instruments ist. In diesem Sinne war AHELO durchaus erfolgreich, indem die grundsätzliche technische, organisatorische und wissenschaftliche Machbarkeit einer internationalen, computerbasierten Kompetenzmessung demonstriert werden konnte. Dass im Ergebnis erhebliche Teile des Messinstruments noch nicht die gewünschte Qualität erreicht haben, die Teilnahmebereitschaft in manchen Länder bedenklich niedrig lag und auch bei der Testdurchführung vor Ort noch manches zu verbessern ist, sollte also nicht als Misserfolg interpretiert werden. Es ist Aufgabe einer Machbarkeitsstudie derartige Probleme zu identifizieren.

Sampling und Feedback

Die unklare Adressatenorientierung hatte im AHELO-Projekt leider auch Auswirkungen auf die Samplingstrategie. Im Hochschulbereich kann ein derartiges Unternehmen nur auf der Basis von Freiwilligkeit durchgeführt werden, das heißt, dass für die Studierenden ein starker Beweggrund für die Teilnahme an einem mehrstündigen Test vorhanden sein muss. Bereits heute zeigen Hochschulrankings, die auf Studierendenbefragungen zurückgreifen die Schwierigkeiten eines solchen Designs. Dort werden meist Rücklaufquoten von unter 20 Prozent erreicht, was bei kleinen Abschlusskohorten schnell zu Auswertungsproblemen führt. Diese Aspekte wurden im AHELO-Projekt stark vernachlässigt, da weder die Integration des Tests in nationale oder lokale Verfahren der Qualitätssicherung hinreichend berücksichtigt, noch die Erfahrungen aus PISA ausreichend ausgewertet wurden und auch eine mögliche Vernetzung von PIAAC oder PISA mit AHELO kaum geprüft wurde. Konkret wäre zu klären, ob neben individuellen Leistungsrückmeldungen an die Testteilnehmer weitere Möglichkeiten der Integration entweder in den hochschulinternen Examinierungsprozess oder in hochschulinterne Evaluationsprozesse realisierbar sind, so dass eine deutlich höherer Beteiligung am Test erreicht werden kann und Verzerrungen durch eine selbstselektive Stichprobe minimiert werden. Ebenso wäre (nicht nur zu Validierungszwecken) zu prüfen, inwieweit etwa aus PIAAC Items übernommen werden könnten, die als Referenz für ein AHELO im Rahmen von Benchmarking genutzt werden könnten. Diese Desiderata sind für eine Weiterentwicklung wichtig, da sie grundsätzlich in eine Richtung deuten, die eher auf ein institutionelles, freiwilliges Benchmarking hinausläuft als auf eine länderrepräsentative Kartierung der weltweiten tertiären Ausbildung. Dabei ist realistischer Weise damit zu rechnen, dass Einrichtungen mit hoher Forschungsreputation an einem solchen Projekt vermutlich nur zögerlich teilnehmen werden, denn bei öffentlicher Zugänglichkeit der Ergebnisse sind die Reputationsgewinne vermutlich gering, die Gefahr potentieller Irritationen durch überraschende Testergebnisse aber groß. Dies wäre aber in einem zukünftigen AHELO durchaus zu verkraften, hätte vielleicht sogar positive Wirkungen, weil es Zeit für inkrementelle Verbesserungen und Raum für Implementationserfahrungen schaffen würde.

Zu bedenkende Folgen: high- or low-stake test?

Assessments wie AHELO lösen regelmäßig eine Fülle von Befürchtungen und Erwartungen, von Widersprüchen und Missbrauchswarnungen und dilemmatischen Entscheidungssituationen aus.

Der Grundkonflikt ist einfach: erhebliche Investitionen müssen gerechtfertigt werden, in der Regel durch den Nachweis von Konsequenzen oder zumindest Verwendungsszenarien. Ebenso beeinflussen die potentiellen Konsequenzen eines Assessments die Teilnahmebereitschaft und Sorgfalt der Testdurchführung. Je gewichtiger aber die Konsequenzen, etwa in Form von Rankings, die Studienwahlentscheidungen beeinflussen, Nutzung in formelbasierten Mittelverteilungssystemen, Integration in Akkreditierungsverfahren, Teilnahmeverpflichtungen durch Integration in Prüfungsverfahren, Belohnung durch ECTS Punkte oder Ähnliches, desto größer die Gefahr strategischen Umgangs, wie „teaching to the test“, missbräuchlicher Verwendung, inadäquater Interpretation der Ergebnisse, Vernachlässigung anderer Leistungsdimensionen der Hochschule, insgesamt konstraintentionaler Nutzung. Umgekehrt gilt: Je folgenloser ein derartiges Assessment ist, desto geringer die Bereitschaft zur Finanzierung, desto problematischer die Mobilisierung der Teilnehmer und in der Folge gilt meist auch, dass die Qualität der Daten unzureichend ist. Diese Debatten begleiteten auch das AHELO Projekt. Wenngleich die Intention, primär zur Qualitätsentwicklung in den Hochschulen beitragen zu wollen, klar formuliert wurde, war das Dilemma dennoch nicht aufzulösen. Beigetragen hat dazu nicht unmaßgeblich die Anlage des Projekts. Manche der teilnehmenden Staaten waren weniger an einer Instrumentenentwicklung als an belastbaren Testergebnissen interessiert, um die nicht unerheblichen Investitionen zu rechtfertigen. Ebenso hat das Ziel möglichst unterschiedliche Zielgruppen zu bedienen, die Phantasien über mögliche Missbrauchsszenarien zusätzlich befeuert. AHELO einfach als low-stake zu definieren wird das Problem nicht lösen, vielmehr bedarf es sehr klarer Regeln, wem welche Informationen in welcher Form zur Verfügung gestellt werden. Deren anschließende Verwendung kann zwar nicht kontrolliert werden, aber Art und Format der Datenbereitstellung können grobem Missbrauch verhindern.

Wie geht es weiter?

Die Finanzierung der AHELO Feasibility Study ist im Dezember 2012 ausgelaufen, faktisch endete AHELO mit einer Abschlusskonferenz im März dieses Jahres. Derzeit ist eine planmäßige Weiterentwicklung nicht im Arbeitsprogramm der OECD verankert. Es hängt also von der Initiative und Finanzierungsbereitschaft einer hinreichend großen Zahl von Staaten ab, ob und wie AHELO weiterentwickelt werden kann. Aus den vorliegenden Ergebnissen ist sehr deutlich erkennbar, dass noch eine Menge Entwicklungsarbeit zu leisten ist. Ebenso klar erkennbar ist aber auch, dass nicht nur bei den Staaten, die an AHELO teilgenommen haben, sondern auch bei vielen, die den Prozess nur beobachtet haben, das Ziel, ein valides und reliables, international einsetzbares Instrument zur Kompetenzmessung in der tertiären Bildung zu entwickeln, sowohl auf Interesse stößt als auch als äußerst relevant eingestuft wird. Insofern geben die Umbruchssituation im AHELO Projekt und die vorliegenden Erfahrungen auch Anlass darüber nachzudenken, wie das Anliegen weiter befördert werden kann. Die Vorteile einer Verortung des Projekts innerhalb der OECD liegen auf der Hand. Ein internationales Assessment bedarf einer Projektsteuerung auf internationaler Ebene. Hier liegen in der OECD nicht nur die Erfahrungen vor, sondern auch viele – teils noch gar nicht aktivierte – Synergien. Organisatorisch und technisch ist daher ein top-down Ansatz – wie er für AHELO gewählt wurde – gar nicht zu umgehen. Die Nachteile sind aber ebenso offenkundig: Die angesichts der enormen Komplexität notwendigen Forschungsarbeiten, die Kontextualisierung innerhalb nationaler Qualitätssicherungssysteme, die Koordination mit standardsetzenden Instituten aus der scientific community und die Verbindung von Kompetenzmessungen mit anderen Leistungsindikatoren (insbesondere der Forschung) sind in einem top-down setting, das stark von der Politik geprägt wird, nur schwer

zu realisieren. Es wäre daher sehr wünschenswert, wenn komplementär ein forschungsorientierter bottom-up Prozess initiiert würde, der sich der vielen im AHELO Abschlussbericht adressierten Fragestellungen annimmt. In Deutschland sind die Voraussetzungen dafür nicht schlecht. Nicht nur Erfahrungen aus der Mitarbeit in internationalen large scale assessments, auch Forschungspotentiale aus einem (gerade auslaufenden) DFG Schwerpunktprogramm zur Kompetenzmessung und der derzeit laufenden BMBF Förderinitiative „Kompetenzmodellierung und Kompetenzerfassung im Hochschulsektor“ bieten eine Ausgangsbasis. Die derzeit starke Orientierung auf Schule, Lehrer und Lehr-Lernprozesse bedürfte allerdings einer deutlichen thematischen Erweiterung auf die spezifischen diagnostischen Anforderungen, wie sie im AHELO Projekt deutlich wurden. Dazu bedürfte es einer gewissen Fokussierung und der Förderung einer deutlich verstärkten Zusammenarbeit zwischen Wissenschaftsforschung und den verschiedenen Feldern der Kompetenzforschung.

Allein Konzept-, Modell- und Instrumentenentwicklung werden aber nicht ausreichen. Kompetenzmessungen enthalten ein unhintergebares normatives Element. Im Falle wissenschaftlicher Ausbildung muss diese Standardsetzung innerhalb der scientific community durch Konsens erreicht werden. Für derartige Setzungen (die zudem mit einem Zeitindex versehen sind) ist die Mitarbeit von Fachgesellschaften und anderen standardsetzenden Instituten der Wissenschaft unerlässlich. Das europäische Tuning Projekt ist ein Schritt in diese Richtung.

Last not least wird es eine zunächst nationale Aufgabe sein, ein Assessment wie AHELO im bestehenden System der Qualitätssicherung zu verorten. Dazu werden Experimentierbereitschaft und breiter Erfahrungsaustausch notwendig sein.

Mit AHELO ist ein in der Zukunft vermutlich äußerst relevant werdender Aspekt der Selbstregulation des Wissenschaftssystems angesprochen worden. Bereits jetzt ist dank des AHELO Projekts erkennbar, dass lediglich die Internationalisierung vorhandener Testkulturen, der Herausforderung nicht gerecht wird. Insofern bietet sich hier eine große Chance an einem Prozess mitzuwirken, der mittelfristig erhebliche Wirkungen auf die Ausgestaltung der tertiären Ausbildung haben wird. Eine Kombination aus nationalen Forschungsstrategien und einer behutsamen Begleitung und Steuerung auf internationaler Ebene wäre daher wünschenswert. Für die deutsche Forschung könnte sich hier die Chance eröffnen aus der Rolle nachholender Entwicklung in die Position eines Schrittmachers zu kommen.

Literatur

- Hornbostel, S.; Lottmann, A. (2012).** Promotionsstrukturen zwischen Autonomie und Koordination. Einleitung zur Studie. In: Hornbostel, S. (Hrsg.): Wer promoviert in Deutschland? Machbarkeitsstudie zur Doktorandenerfassung und Qualitätssicherung von Promotionen an deutschen Hochschulen. iFQ-Working Paper No. 14. Berlin, 9–17.
- OECD (2011).** AHELO: Assessment of Higher Education Learning Outcomes. Paris. Online-Dokument: <http://www.oecd.org/edu/skills-beyond-school/45755875.pdf> (15.03.2013)
- OECD (2012).** Assessment of Higher Education Learning Outcomes. Feasibility Study Report. Volume 1 – Design and Implementation. Paris. Online-Dokument: <http://www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume1.pdf> (15.03.2013)
- OECD (2013).** Assessment of Higher Education Learning Outcomes. Feasibility Study Report. Volume 2 – Data Analysis and National Experiences. Paris. Online-Dokument: www.oecd.org/edu/skills-beyond-school/AHELOFSReportVolume2.pdf (15.03.2013)

Osterberg, K.; Kölbl, S.; Brauns K. (2006). Der Progress Test Medizin: Erfahrungen an der Charité Berlin. *GMS Z Med Ausbild.*, 23(3), Doc46.

Wissenschaftsrat (2012). Prüfungsnoten an Hochschulen im Prüfungsjahr 2010. Arbeitsbericht mit einem Wissenschaftspolitischen Kommentar. Drs. 2627-12. Hamburg. Online-Dokument: <http://www.wissenschaftsrat.de/download/archiv/2627-12.pdf> (15.03.2013)

AHELO – Zwischen Wissenschaft und Politik

Hans Pechar

Im vorliegenden Beitrag soll in drei Komplexen auf das Spannungsfeld zwischen Wissenschaft und Politik, in dem sich das OECD-Projekt AHELO befindet, eingegangen werden. Zunächst werden aktuelle Problemlagen, die sich aus ersten Erkenntnissen der Pilotstudie von AHELO ergeben, kurz erläutert. Daran anschließend wird diskutiert, ob AHELO als eine Art „PISA für die Hochschulen“ zu betrachten ist. Abschließend wird die Position der österreichischen Wissenschaftspolitik vorgestellt. Meine Kernthese ist, dass internationale Vergleichsstudien, die Kennwerte aus Bildungssystemen berichten, sich immer in der schwierigen Situation befinden, sowohl politische als auch wissenschaftliche Projekte zu sein. Für die Bildungspolitik sind oftmals weniger wissenschaftliche und methodologische Überlegungen bei ihrer Entscheidung für oder wider ein solches Assessment leitend als vielmehr Fragen nach der Wirkung der Ergebnisse in der Öffentlichkeit. So muss man die Gefahren einer politischen Instrumentalisierung immer bedenken.

Die AHELO-Machbarkeitsstudie wird seit 2008 durchgeführt und umfasst die Bereiche der *generic skills*, also der überfachlichen Kompetenzen, sowie die *discipline-specific skills* in den Fächergruppen der Ingenieurs- und Wirtschaftswissenschaften.

Eigentlich sollte in 2012 bereits eine Entscheidung zur Zukunft von AHELO fallen, sollte die Pilotstudie soweit ausgewertet sein, dass man auf den Erfahrungen basierend über Modifikationen an AHELO und eine Fortführung mit ausgedehntem Teilnehmerkreis nachdenken konnte. Doch diese Entscheidungen sind auf den März 2013 vertagt. Es wird aber bereits jetzt eine Reihe von kritischen Punkten deutlich, die die Pilotstudie sichtbar gemacht hat und die im Vorfeld möglicher Entscheidungen über AHELO diskutiert werden:

- Wer entscheidet letztlich über die Teilnahme an AHELO: Die jeweilig zuständige nationale oder föderale Regierung? Oder sollten die Hochschulen eigenständig über die Teilnahme entscheiden dürfen?
- Was wird eigentlich verglichen? Die Kompetenzen von je individuellen Studierenden oder die Leistungsfähigkeit von Hochschulen bei der Kompetenzvermittlung?
- Ist beabsichtigt den value-added, also den spezifischen durch das Studium und die Lernumwelt verursachten Kompetenzzuwachs auszuweisen? Ist es (in organisatorischer und finanzieller Hinsicht) möglich, valide Vorher- / Nachhermessungen durchzuführen?
- Wie gewichtet man abschließend das Verhältnis von generischen und fachspezifischen Kompetenzen? Welche Fachunterschiede müssen dabei berücksichtigt werden?
- Und wie kann sich AHELO dauerhaft finanzieren?
- AHELO muss zudem in den Kontext ähnlich gelagerter Forschungsprojekte gestellt werden. So werden im OECD-Projekt PIAAC (Programme for the International Assessment of Adult Competencies) aktuell jene Kompetenzen von 16- bis 65-Jährigen im internationalen Vergleich untersucht, die für eine Teilhabe am gesellschaftlichen Leben sowie Erfolg im Beschäftigungssystem notwendig sind. Wie ist AHELO hier anschlussfähig? Wie können Überschneidungen zwischen den Studien vermieden werden?

An der AHELO-Pilotstudie nehmen höchst unterschiedliche Länder teil: Ägypten, Australien, Belgien, Finnland, Italien, Japan, Kolumbien, Kuwait, Mexiko, Niederlande, Norwegen, Russland,

Schweden und die USA. Auffallend ist, dass sehr viele der teilnehmenden Länder keine OECD Mitglieder sind. Möglicherweise hängt das damit zusammen, dass AHELO bei den Kernmitgliedern der OECD noch nicht auf ausreichende Resonanz stößt. Aber wie wird sich die starke Heterogenität auf die Vergleichbarkeit der Ergebnisse auswirken? Zweitens fällt auf, dass sich innerhalb der OECD „die üblichen Verdächtigen“ beteiligen, also vorwiegend Länder, die an den meisten Vergleichsstudien der OECD teilnehmen, während Süd-, Ost- und Mitteleuropa fast nicht vertreten ist.

Interessant ist auch, dass in den USA erst die Regierung Obama Interesse für AHELO gezeigt hat, während die Bush-Administration, die im Übrigen sehr stark auf „high stake testing“ gesetzt hat, sich nicht beteiligen wollte. Die Erhebung der generischen Kompetenzen erfolgt in der Pilotstudie durch eine Adaption des US Collegiate Learning Assessment (CLA). Das CLA testet die Kompetenz von Studierenden in den Bereichen logisches Denken, kritisches Denken, Problemlösungsfähigkeiten und sprachliche Ausdrucksfähigkeit (schriftlich). Der Test ist open-ended und low-stake angelegt, d. h. er wird ohne vorgegebene Antwortmöglichkeiten und ohne starke formale Einbindung ins Studium durchgeführt. Und schließlich ist CLA ein summativer Test, der nicht nur Kompetenzniveaus, sondern auch Kompetenzgewinne während des Studiums erfasst.

AHELO wird des Öfteren als *PISA für die Hochschule* bezeichnet. Doch welche Berechtigung hat diese Analogie? Zum einen gibt es sicherlich Gemeinsamkeiten zwischen PISA und AHELO, stehen doch beide Assessments für zwei wichtige globale Trends im Bildungsmonitoring: (1) Mit internationalen Vergleichen geht eine gewisse Unabhängigkeit von nationalen Zertifizierungsagenturen und den dahinter liegenden je nationalen Vorstellungen von „guter Bildung“ einher. Es sind nicht mehr lokale Autoritäten, die die Definitionsmacht für die Qualität der Bildungsinstitutionen innehaben, vielmehr verständigen sich globale Eliten auf Standards. (2) Der formale Abschluss als Signal für Qualifikationen verliert an Bedeutung, dagegen rücken empirisch messbare, im Studium erworbene Kompetenzen zunehmend in den Mittelpunkt. Zum anderen finden sich auch Unterschiede zwischen den beiden Assessments. Während bei PISA eine nationale Stichprobe gezogen wird, erfolgt der Prozess der Auswahl bei AHELO selbstselektiv, d. h. einzelne Hochschulen erklären ihre Teilnahme. Eine repräsentative Untersuchung nationaler Hochschulsysteme kann so nicht geleistet werden. Bei AHELO liegt der Fokus der Untersuchung auf dem Lernzuwachs, bei PISA auf dem Lernstand. Dies bedingt je unterschiedliche methodische Herangehensweisen und führt zu einer unterschiedlichen Reichweite und Gültigkeit von Interpretationen. So kann PISA eher eine Momentaufnahme des Wissens von Schüler(innen) liefern, während AHELO die Kompetenzgewinne von Studierenden durch das Absolvieren ihres Hochschulstudiums sichtbar machen möchte, auch wenn dies derzeit noch nicht umgesetzt wird. Und schließlich wird aktuell darüber diskutiert, ob AHELO ein Ranking ist resp. so verwendet wird (s. a. die Beiträge von Barbara M. Kehm und Susanne in der Smitten in diesem Band). Problematisch im Umfeld von PISA ist die medienwirksame Veröffentlichung der Ergebnisse, die letztlich in der Selektions- und Präsentationslogik der Massenmedien zur bloßen Reihung von Staaten entlang ihres Abschneidens führt und PISA damit auf ein Ranking reduziert. Solche Interpretationen aber bergen wiederum politische Sprengkraft, könnte die Reaktion auf eine vermeintlich schlechte Platzierung bei Assessments wie PISA oder AHELO zu kurzfristigen, symbolischen Interventionen der Bildungspolitik oder gar zu einem Ausstieg aus dem Projekt führen.

Abschließend möchte ich daher noch auf die Position der österreichischen Politik zu AHELO eingehen. Bislang lehnt das zuständige Ministerium eine Teilnahme an AHELO ab, ein Richtungswechsel oder auch nur eine Diskussion über dieses Thema zeichnet sich nicht ab. Generell gibt es bei maßgeblichen Akteuren der österreichischen Bildungspolitik eine starke Skepsis gegenüber Standards im Bildungsbereich – was beispielsweise in der Verschiebung des Projekts der Zentral-

matura zum Ausdruck kommt. Die Lehrer/innengewerkschaft fordert sogar einen Ausstieg aus PISA – immerhin hat sie einen Ausstieg aus TALIS (Teaching and Learning International Survey) durchgesetzt. Damit zeigt sich beispielhaft, dass die politischen Entscheidungen über ein solches Assessment sich nicht an den wissenschaftlichen Diskussionen, die dieses begleiten, und den daraus abgeleiteten Forschungsbedarfen orientieren. Der Erfolg von international vergleichenden Studien der empirischen Bildungsforschung bemisst sich demnach nicht rein an der wissenschaftlichen Validität der Instrumente oder dem Schließen von Forschungslücken. Ihr Erfolg hängt auch davon ab, wie die Analysen im politischen System und der Öffentlichkeit interpretiert und verarbeitet werden – und darauf hat die Wissenschaft m. E. wenig Einfluss.

Kompetenzmessung im Hochschulbereich. Wege aus der Blackbox – Ein Interview mit Stefan Hornbostel und Edith Braun

Tanja Meister & André Donk

Im Anschluss an den in diesem Band dokumentierten AHELO-Workshop der Gesellschaft für Hochschulforschung traf das HIS-Magazin Dr. habil. Edith Braun, Wissenschaftliche Leiterin des HIS-Instituts für Hochschulforschung, und Prof. Dr. Stefan Hornbostel, Leiter des Instituts für Forschungsinformation und Qualitätssicherung (iFQ), zum Fachgespräch über Kompetenzmessung im Hochschulbereich. Eine Kurzfassung des Gesprächs ist im HIS-Magazin 1|2013 (S. 3-5) erschienen; die Langfassung wird erstmals an dieser Stelle abgedruckt.

Was verstehen Sie unter Kompetenzen?

Braun: „Kompetenzen sind erlernbar. Das heißt: Es handelt sich hierbei nicht um ein Merkmal, das angeboren und stabil ist, sondern das über Bildungsprozesse erworben wird und daher veränderbar, erweiterbar, förderbar ist. Kompetenzen sind also Ergebnis eines Bildungsprozesses. Wichtig ist, dass Kompetenzen immer situationsabhängig sind, also spezifisch für einen bestimmten Kontext definiert werden können. Im Hochschulbereich reden wir nicht nur über fachspezifische Kompetenzen, sondern auch über überfachliche Kompetenzen. Es kommen damit ganz wichtige motivationale, soziale und persönliche Merkmale zum Tragen.“

Hornbostel: „Ich kann eigentlich nur noch ergänzen, dass hier auch Wissen eine Rolle spielt. Wissen und Kompetenzen sind allerdings nicht identisch. Es geht bei Kompetenzen in der Regel darum, Erlerntes in bestimmten Situationen anwendbar zu machen.“

Wie kann man Kompetenzen – derart definiert – messbar machen? Kann man Kompetenzen überhaupt messen?

Braun: „In der Psychologie haben wir häufig den Fall, dass wir Konstrukte messen, die im ersten Augenblick nicht messbar erscheinen. Das Wichtigste ist zu definieren, worüber wir sprechen, also uns auf eine theoretische Bezeichnung dessen, was wir eigentlich messen wollen, zu einigen. In Bezug auf Kompetenzen habe ich meine Zweifel, ob wir diese ausschließlich über schriftliche Leistungstests messen können oder ob wir nicht auch andere Zugänge brauchen, wie z. B. *computer-based testing*, wo wir Situationen schaffen, in denen wir nicht nur das reine Fachwissen, sondern auch – wie von Herrn Hornbostel eben erläutert – Handlungsfähigkeit abtesten.“

Hornbostel: „„Messen“ ist ein Begriff, bei dem man sofort an Zentimetermaß denkt. Stellen wir uns einmal ganz lebensweltlich eine Situation vor, in der man eine Dienstleistung in Anspruch nimmt. Sie bestellen also zum Beispiel einen Handwerker, weil Ihr Wasserhahn kaputt ist, und vertrauen darauf, dass er sein Fachwissen in einer komplexen Situation so anwenden kann, dass er Ihren Wasserhahn wieder zum Laufen bringt. Sie können dann auch beurteilen, ob das geklappt hat und ob Sie mit der Leistung zufrieden waren. Man hat also ein Gespür dafür, ob jemand in einer bestimmten Situation kompetent handelt oder nicht. Um das messbar zu machen, kann man sich ein ganzes Spektrum von Methoden vorstellen, aber vom Ansatz her ist erst einmal klar: Wenn man selbst eine Idee davon hat, dass es kompetentes Handeln gibt, dann gibt es

auch einen Weg, das intersubjektiv nachvollziehbar zu machen. Welche Methoden anschließend zum Einsatz kommen, ist im Grunde genommen eine technische Frage. Da wird man viel experimentieren müssen, bis man ein Verfahren hat, von dem man einigermaßen sicher sagen kann: Es misst die gewünschten Konstrukte tatsächlich.

Welche Kompetenzen können Studierende in einem Hochschulstudium eigentlich erwerben? Und welche sollten sie erwerben?

Hornbostel: „Das ist sehr stark in Bewegung gekommen. Noch vor 40 Jahren gab es eine simple Anforderung: Man lernt den akademischen Habitus und wenn man den einigermaßen stolperfrei beherrscht, ist das Ziel eines Hochschulstudiums erreicht. In der Zwischenzeit – das ist insbesondere am Bologna-Prozess ablesbar – ist eine Fülle von Anforderungen und Erwartungen hinzugekommen, welche Kompetenzen im Rahmen einer Hochschulausbildung erworben werden sollen. Das reicht vom klassischen Fachwissen über das Nachholen von bestimmten Sozialisationsdefiziten bis hin zu Anforderungen aus den Arbeitsmärkten und übergeordneten bzw. extrafunktionalen Qualifikationen und Kompetenzen. Man kann zu Recht fragen, ob das alles eigentlich noch zusammengeht. Zumindest im deutschen Hochschulsystem stoßen an dieser Stelle zwei Traditionen aufeinander. Ein Teil dieser Anforderungen kommt aus der britischen Gentleman-Erziehung, wo Aufgabe der Hochschulen die Elitenerziehung war. Hier erhielten die jungen Leute den letzten Schliff. Im deutschen Hochschulsystem ist das anders: Wir haben seit der Humboldt'schen Reform eine stark fachwissenschaftliche Ausbildung. Viele der extrafunktionalen Qualifikationen entstammen aber eher der Idee der erziehenden Institution, die Menschen formt. Die deutsche Tradition geht davon aus, dass diese überfachlichen Qualifikationen von den Studierenden mitgebracht oder außerhalb der Hochschule selbst erworben werden. Vielleicht ist das ein Grund, warum wir uns in Deutschland besonders schwer mit der Formulierung von Anforderungen an die hochschulische Ausbildung tun.“

Ein Instrument zur Messung hochschulischer Kompetenzen muss – nach allem, was wir gehört haben – vermutlich sehr komplex und aufwendig konstruiert sein, zumal wenn es auch noch international zum Einsatz kommen soll. Reichen bestehende Instrumente, wie z. B. Rankings, denn nicht aus? Worin liegt der Mehrwert eines Instruments zur Kompetenzmessung?

Braun: „Eine Kompetenzmessung hilft, auf verschiedensten Ebenen zu steuern. Zum Beispiel auf der Ebene der Institution: Dadurch dass man deutlich macht, was die Ergebnisse hochschulischer Bildung sind, und sie prüfbar und messbar macht, steuert man auch die Lehrenden in der jeweiligen Fakultät. Das ist wichtig, wenn man z. B. über Qualitätssicherung spricht. Ein Vorteil ist sicher der Blick auf das Lernergebnis, darauf, was man mit Bildung erreicht.“

Hornbostel: „Vielleicht hilft auch hier ein Rückblick. Um die vorletzte Jahrtausendwende waren die maßgeblichen Qualitätselemente eines Hochschulstudiums, an welcher Universität und bei wem man studiert hat. Und damit war eigentlich die Kompetenzfrage auch schon beantwortet. Wenn man große Geister aufzählen konnte, bei denen man im Hörsaal gesessen hatte, war klar: Man hat an dem Heiligenschein dieser Genies in irgendeiner Form partizipiert. Nach Einsetzen der ersten Vermassungstendenzen nach dem Zweiten Weltkrieg kam etwas anderes ins Spiel: die Benotung. Sie teilte mit, ob jemand besonders gut oder schlecht war. Heute ist es schwierig, Noten noch eine Aussagefähigkeit zu attestieren. Sie schnurren im Grunde zu Einheitsnoten zusammen. Es gibt eine Inflation von guten Noten und wenig Differenzierung. Hinzu kommt die Europäisierung des Hochschulraums. Die Frage ist jetzt: Was kann ich konkret von jemandem, der einen

hochspezialisierten Bachelor oder Master an einer spanischen Universität erworben hat, erwarten? Mit der Note kann ich nur noch wenig anfangen, weil ich nicht weiß, was damit wirklich zertifiziert wird. An ganz verschiedenen Stellen entsteht also ein Bedarf zu wissen, was ein Absolvent einer Einrichtung im Schnitt kann. Und das ist nicht nur für die Abnehmer, also den Arbeitsmarkt, wichtig, sondern auch für die Universitäten – als Rückmeldemechanismus, wo sie im internationalen Vergleich stehen. Eine solche Rückmeldung ist last but not least auch für die Studierenden selbst wichtig, die sonst immer erst ex post, nämlich aus der Berufssituation heraus, sagen können, ob in der hochschulischen Ausbildung etwas gefehlt hat bzw. wo die eigene Qualifikation besonders gut ist. Auch für die Studierenden wäre also eine Rückmeldung darüber, was sie selbst erfüllen, wünschenswert und hilfreich. Wir haben somit ganz verschiedene Adressatengruppen und Zielsetzungen für eine Kompetenzmessung. Für die einen dient sie der Steuerung, für die anderen der Selbstverortung, für Dritte birgt sie Reflexionspotenzial.

Braun: ¶¶ Diesen Gedanken möchte ich gerne aufnehmen. Genau das zeichnet doch die Wissenschaft aus: Sie bewegt sich schon immer in einer wettbewerbsorientierten Landschaft. Wir bekommen über Reviewprozesse sehr deutlich rückgemeldet, wie unsere Publikationen von unseren Peers eingeschätzt werden. Man wird gemessen an der Anzahl der Publikationen in referierten Fachzeitschriften und den eingeworbenen Drittmitteln. Es handelt sich um ein selbstgesteuertes System, das davon lebt, dass sich Peers gegenseitig Rückmeldung geben. Hier sind wir sehr nahe an dem, was Wissenschaft im Kern ausmacht, nämlich dass man sich durch die Rückmeldung einordnet und weiterentwickelt und das nicht als Gefahr sieht, sondern für sich nutzt.

Hornbostel: ¶¶ Wobei Ihr Beispiel auch die dunklen Seiten zeigt. Gerade im Bereich des Peer Review haben wir auch eine ganze Menge Probleme. Auch hier stellt sich die Frage: Wie misst man die Qualität von Wissenschaftlern? Schaut man auf den Impact Faktor des Journals? Bei den Manuskriptbegutachtungen durch Peers gibt es die Kritik, sie seien sehr konservativ, risikoscheu und mainstream-orientiert. Solche nicht-intendierten negativen Effekte könnte man sich natürlich auch im Bereich der Kompetenzmessung vorstellen. All das muss man mit bedenken, wenn man entsprechende Instrumente entwickelt.

Welche Informationen liegen gegenwärtig zu Fragen der Kompetenzmessung und des Kompetenzerwerbs im Hochschulbereich in Deutschland vor?

Braun: ¶¶ Hier gibt es eine deutliche und berechtigte Kritik an der Hochschulforschung und auch an der Bildungsforschung insgesamt. Wir haben eine ganze Reihe von Instrumenten zur Erfassung von Selbsteinschätzungen. Für Forschungszwecke sind Selbsteinschätzungen hochrelevant und interessant. Das hat sich immer wieder gezeigt. Aber wir sind uns einig, dass sie für die Diagnostik – erst recht auf Individualebene – nicht genutzt werden können. Instrumente zur Messung von Kompetenzen oder gar von Kompetenzentwicklung – dem sog. *value added* bzw. *gain of competences* – fehlen uns allerdings in weiten Teilen. Wir machen erste Erfahrungen im internationalen Vergleich mit der AHELO-Studie. In Deutschland gibt es eine Förderlinie des BMBF zur Kompetenzmessung im Hochschulbereich. Vor allem in den Lehramtsstudiengängen ist man hier bereits ein Stück vorangekommen. Insgesamt besteht in der Hochschulforschung in meinen Augen aber ein ganz starker Forschungsbedarf zu Kompetenzerwerb und Kompetenzentwicklung.

Hornbostel: ¶¶ Auch hier können Sie das Problem von hinten aufrollen: Warum haben wir eigentlich bis heute Staatsprüfungen? Die haben wir genau in den Bereichen – wie etwa Medizin, Jura

oder Lehramt –, in denen der Staat ein massives Interesse an Qualitätssicherung hat. Im Gesundheitsbereich ist es zum Beispiel ganz wichtig, dass man garantieren kann, dass die Absolventen bestimmte Qualifikationen erworben haben. Hier signalisiert der Staat ein erhebliches Misstrauen in die Fähigkeit der Hochschulen, Qualität zu sichern. Nun sind staatliche Prüfungen kein elaboriertes Instrument, um Qualität zu messen. Aber der Staat versucht mit eigenen Regularien, Anforderungen an die Testsituation zu formulieren und zu definieren, was ein Absolvent können soll. Was darüber hinaus das Wissen über Kompetenzen angeht, tappen wir tatsächlich über weite Strecken im Dunkeln. Es ist daher sicher kein Zufall, dass Projekte wie AHELO vor dem Hintergrund steigender Studierendenzahlen und einer starken Diversifizierung der Abgangsqualifikationen gerade jetzt starten.

Braun: „Ich würde gerne einen Gedanken aufnehmen: Das Beispiel der Staatsprüfungen ist ein gutes Beispiel. Hierbei handelt es sich nämlich tatsächlich um kompetenzorientiertes Prüfen. Ich glaube allerdings nicht, dass Staatsprüfungen Ausdruck von Misstrauen gegenüber den Hochschulen sind, sondern dass dahinter die Erkenntnis steht, dass zentrale Standards der Qualitätssicherung dienen. Von außen zentrale Kriterien vorzugeben, nach denen Qualität gemessen wird, und dies nicht dem System selbst zu überlassen, ist das, was wir kriterienorientiertes Prüfen oder Messen nennen.

Hornbostel: „Historisch betrachtet spielen sicher drei verschiedene Ebenen eine Rolle: Was sie ansprechen – die Selbstreferenzialität von Wissenschaft – ist typisch für den Forschungsprozess. Forschung definiert sich so, dass sie ihre Bewertungskriterien aus sich selbst heraus gewinnt. Es gibt wenige Kriterien von außen, die definieren, ob eine Forschung gut, richtig und relevant ist. Was hingegen im 18. und 19. Jahrhundert an den Universitäten passierte, ist ausgesprochen bedenklich: Die Promotionen wurden verkauft, denn sie brachten Geld. Die Folge war ein Wettbewerb, der sich im Prinzip darin erschöpfte, die Qualität immer weiter zu senken. Die Fürstenhöfe reagierten auf diese Entwicklung mit der Einführung klarer – häufig sogar überregionaler – Prüfungsstandards. Hier wurde der selbstreferenzielle Zirkel verlassen. Stattdessen wurden Anforderungen aus der Praxis definiert, die von den Absolventen erfüllt werden mussten. Der Weg dorthin blieb der Universität selbst überlassen. Der Prüfer will am Ende nur wissen: Kann der Prüfling die Situation, vor die er gestellt wird, erfolgreich bewältigen?

In der Kompetenzmessung liegt in Deutschland – das haben Sie beide betont – zumindest im Hochschulbereich vieles noch im Dunkeln. Zugleich haben wir heute Morgen über eine konkrete Initiative gesprochen – das AHELO-Projekt der OECD. Ist AHELO die Lösung?

Hornbostel: „Die Lösung ist es sicherlich nicht. AHELO ist ein spannender Versuch, mit dem in vielerlei Hinsicht Neuland betreten wird. Hier sollen Kompetenzen unabhängig vom kulturellen Hintergrund gemessen werden. Das ist an sich schon eine hohe Anforderung. Es geht zudem um viel abstraktere Fähigkeiten als z. B. bei PISA. Da spielen dann sehr schnell normative Aspekte eine Rolle. Ob Kritikfähigkeit wirklich ein universales Merkmal akademischer Bildung ist, wird in Asien vermutlich anders beurteilt als in Kontinentaleuropa oder den USA. Hinzu kommt, dass die Frage, was genau ein Ingenieur, ein Sozialwissenschaftler, ein Lehrer oder ein Mediziner können muss, keineswegs abschließend geklärt ist. Last but not least greifen Sie mit einer solchen Definition von Kompetenzen – und deshalb ist sie auch so umstritten – stark in die akademische Autonomie ein. AHELO ist also ein interessanter erster Schritt und man wird viel lernen können, was die internationale Vergleichbarkeit und geeignete Verfahren betrifft.

Braun: „ Ganz sicher ist AHELO ein erster Schritt, sich der Black Box Kompetenzmessung zu nähern. Welche Merkmale der Lehrgestaltung oder auch der Lehrperson wirken sich eigentlich wie auf das Lernen der Studierenden aus? Im Hochschulbereich wissen wir hierüber nichts. In der Schulforschung wissen wir hier schon deutlich mehr. Natürlich haben wir in der Hochschulforschung ganz andere methodische Herausforderungen als in der Schulforschung. Es gibt z. B. kein einheitliches Curriculum, sondern die Studierenden besuchen Lehrveranstaltungen und Module, die sich völlig unterschiedlich zusammensetzen. Sie besuchen Veranstaltungen verschiedener Lehrpersonen. AHELO kann dazu beitragen, dass wir uns darüber klar werden, wie wir Kompetenzen operationalisieren und messen wollen. Erst wenn wir solche Messinstrumente haben, können wir diese in Verbindung mit Merkmalen der Lehrgestaltung bringen. Der ganze Bereich der Lehr-/Lernforschung in der Hochschule ist bislang eine große Black Box.

Hornbostel: „ AHELO eröffnet noch eine andere Möglichkeit, nämlich sich ganz von diesen curricularen Bedingungen zu trennen und wirklich nur auf den Output bzw. Outcome zu schauen, ohne zu berücksichtigen, wie dieser zustande gekommen ist. Dann wäre es unerheblich, ob ein Student die Lehrveranstaltungen besucht oder stattdessen nur das Begleitbuch gelesen und anschließend die Prüfung erfolgreich abgelegt hat.

Braun: „ Ich gebe Ihnen Recht. Als Psychologin würde ich dennoch sagen, dass wir systematische Einflüsse untersuchen. Das ist es, was uns in der Lehr-/Lernforschung interessiert. Was sind die Hebel, an denen wir ansetzen können, um die Qualität in der Bildung zu steigern? Das können wir nur erforschen, wenn wir Kompetenzen erfassbar machen und dann den Zusammenhang zur Lernumgebung untersuchen. Es ist natürlich richtig, dass Studierende sehr individuelle Bildungswege durchlaufen; wenn wir aber systematische Zusammenhänge finden, können wir Rückschlüsse auf die Lehrqualität ziehen.

Hornbostel: „ Da stimme ich Ihnen völlig zu. Das Wissen um solche Kompetenzstrukturen ist im Grunde genommen der Motor für curriculare Reformen. Wenn man kein Signal hat, dass etwas defizitär ist, gibt es auch keinen Grund, etwas zu ändern. Es gibt eine Reihe solcher Signale, z. B. Rückmeldungen von den Arbeitgebern über Kompetenzdefizite von Absolventen oder hohe Dropout-Quoten, aber keine systematische Forschung.

Ist es denn überhaupt möglich sicherzustellen, dass wir in unterschiedlichen sprachlichen und kulturellen Kontexten das Gleiche messen?

Hornbostel: „ Aus der AHELO-Studie kann man durchaus ableiten, dass das machbar ist. Es ist aufwendig und ein langer Prozess mit vielen Stolpersteinen, bis ein solches Instrument implementiert ist, aber es geht.

Braun: „ Wir können hier viel aus anderen Testbereichen lernen. Intelligenzforschung z. B. hat eine lange Tradition und auch hier hat man internationale Tests entwickelt, die über kulturelle Grenzen hinweg einsetzbar sind. An der Intelligenzforschung kann man allerdings auch ablesen, dass eine Testentwicklung immer normativ ist. Im Gegensatz zu früher teilt man heute z. B. die Überzeugung, dass Frauen und Männer oder auch verschiedene ethnische Gruppen in Intelligenztests gleich abschneiden sollten; dementsprechend hat man das Testdesign angepasst. Wie wir unsere Instrumente anpassen und weiterentwickeln müssen, können wir aber nur über den

internationalen Vergleich lernen. Die Erfahrung in der Testentwicklung müsste heute eigentlich dazu führen, dass ein Test von vornherein so konstruiert wird, dass er international einsetzbar ist.

Hornbostel: „ Da sprechen Sie wahrscheinlich gerade das heikelste Thema im Testbereich an. Wenn wir einmal hypothetisch davon ausgehen, es gäbe tatsächlich Unterschiede in der Intelligenzverteilung zwischen Männern und Frauen oder zwischen ethnischen Gruppen, Sie aber von der normativen Prämisse ausgehen, es gibt sie nicht und entsprechend Ihre Items anpassen ...

Braun: „ Aber so wird doch vorgegangen! Eine Testentwicklung ist immer normativ. Sie ist ein bewusster Entscheidungsprozess, welche Fragen gestellt werden und welche nicht.

Hornbostel: „ Klar, das stimmt absolut. Und diese Normativität muss nicht unbedingt positiv sein. Deshalb muss man das bewusst reflektieren.

Braun: „ Ganz genau! Man muss ein Bewusstsein dafür haben und eine Einigung darüber herstellen, was ein Test leisten soll und was nicht. Aber auch da sehe ich eine Chance: Deutschland ist prädestiniert dafür, seine Erfahrungen mit der Testentwicklung auch international einzubringen.

Die Testzeit für die Studierenden beträgt bei AHELO zweieinhalb Stunden. Wie bringt man die Studierenden dazu, sich an einem solch aufwendigen Test zu beteiligen? Und wie kann man auch die „erhebungsmüden“ Hochschulen davon überzeugen, sich einzubringen?

Braun: „ Da wäre ich so mutig und würde einfach einmal ausprobieren, das als Lehrveranstaltung durchzuführen. Diesbezüglich fand ich einen Vorschlag von heute Morgen ganz hervorragend: Man sucht sich drei Hochschulen für eine solche Testphase und integriert den Test in die Lehrveranstaltungen. Das machen wir in anderen Bereichen doch auch. Für die Teilnahme bekommen die Studierenden anschließend Kreditpunkte.

Hornbostel: „ Da muss man dann aber die Testzeit anpassen! Eine Lehrveranstaltung dauert nur zweimal fünfundvierzig Minuten. Danach sind die Studierenden weg. (lacht) In der Tat ist die Teilnahmebereitschaft ein großes Problem. Anders als bei den Schultests gibt es kaum eine Möglichkeit, die Teilnahme verbindlich zu machen. Universität ist und bleibt eine freiwillige Veranstaltung. Die Frage ist im Grunde genommen, welche Incentives man entwickeln kann, um Studierende zur Teilnahme zu bewegen. Interessant ist, dass in Asien und in den Schwellenländern die Teilnahmebereitschaft deutlich höher liegt. Das deutet darauf hin, dass es bei den Studierenden entweder ein Verpflichtungsgefühl gibt, oder es ist die Neugier, etwas darüber zu erfahren, was das für Aufgaben sind und ob ich als Student z. B. in Bangladesch damit klarkomme. Das kann gerade für diejenigen interessant sein, der plant, zu einem späteren Zeitpunkt z. B. nach Europa oder in die USA zu gehen. In den europäischen Ländern ist die Frage der Teilnahmebereitschaft weitaus komplizierter. Die Neugier ist geringer, an ihre Stelle tritt eher ein Genervtsein – schon wieder ein Fragebogen bzw. Test! Klar, man kann mit ECTS-Punkten locken. Man kommt dann allerdings in eine Situation, wo man den Test als Lehrveranstaltung verkaufen muss. Das würde eigentlich voraussetzen, dass man dabei etwas lernt...

Braun: „ Aber das tritt ja ein! Es wird ein Selbstreflexionsprozess ausgelöst, wie von Ihnen eben angesprochen.

Hornbostel: ¶¶ Das ist natürlich nicht ausgeschlossen. Wir bekommen allerdings ein Problem, was spätere statistische Auswertungen betrifft. Wir haben immer eine Selbstselektivität in der Teilnehmerstruktur. Wir wissen nicht, ob diejenigen, die genervt sind und nicht teilnehmen, und diejenigen, die neugierig sind und teilnehmen, sich nicht auch in ganz vielen anderen Aspekten unterscheiden. Man muss da, glaube ich, auch als Hochschule ein bisschen Phantasie entwickeln. Es wird nicht reichen, nur einen netten Brief des Präsidenten mit der Aufforderung zur Teilnahme zu verschicken.

Braun: ¶¶ Ich möchte unterstreichen, dass selbst dann, wenn man den Test im Rahmen einer Lehrveranstaltung durchführen würde, dies immer nur auf freiwilliger Basis geschehen kann. Was aus einer Testentwicklungsperspektive interessant ist, ist aber gerade diese Streuung, die Sie angesprochen haben, dass man die Gelangweilten dabei hat, dass man die Überflieger dabei hat. Wir brauchen die ganze Bandbreite.

Diese Wende hin zu einer Fokussierung auf Kompetenzen – was bedeutet sie für die Hochschulforschung in Deutschland, was für die Wissenschaftsforschung?

Braun: ¶¶ Ich würde die Hochschulforschung zu weiten Teilen der empirischen Bildungsforschung zuordnen. In der empirischen Bildungsforschung ist es eine zentrale Frage, welche Kompetenzen Studierende erwerben sollen, wie sie ihre Kompetenzen entwickeln und wie wir als Lehrende sie darin unterstützen können. Welche Merkmale der Lehrpersonen sind hilfreich in diesem Prozess?

Sehen Sie die deutsche Hochschulforschung hier gut aufgestellt?

Braun: ¶¶ Der heutige Tag hat eigentlich wieder gezeigt, dass wir hier ganz am Anfang stehen. Es gibt in der deutschen Hochschulforschungslandschaft keine institutionelle Verortung dieser Fragen.

Hornbostel: ¶¶ Ich würde hier keine Trennungslinie zwischen Hochschulforschung und Wissenschaftsforschung ziehen. Wir diskutieren seit 20 Jahren, inwieweit Forschung gesellschaftliche Probleme aufnehmen soll und wie man diese in die Forschung transportieren kann. Kurz: Wie macht man Forschung nützlich? Relativ klar ist, dass eine moderne Gesellschaft eine starke Basierung in der Wissenschaft und im Wissen hat und darauf angewiesen ist, dass es ihr gelingt, ihre großen Probleme in Forschung und Wissenschaft zu transportieren und sie dort bearbeitbar zu machen. Im Grunde genommen stellt sich genau die gleiche Frage auf der Ausbildungsseite. Es reicht nicht, nur in die Autopoiesis der Wissenschaft eingeführt zu werden, sondern es ist immer wichtiger geworden, dieses Wissen im Berufskontext, aber auch in anderen gesellschaftlichen Problemkontexten, anwenden zu können. Phänomenologisch ist beides sehr ähnlich: Die Forschung dockt an gesellschaftliche Probleme an und macht diese bearbeitbar, ohne die Selbstreferenzialität als Grundmodus von Forschung zu verlieren. In der Ausbildung stellt sich die Frage: Wie gelingt es, eine autonome wissenschaftliche Ausbildung zu garantieren, die aber gleichzeitig die Kompetenzen zum Meistern praktischer Verwendungssituationen vermittelt? Wir haben also ganz ähnliche Probleme in Forschung und Ausbildung.

Herzlichen Dank für das ausführliche Gespräch!

Ergebnisprotokoll zum AHELO-Workshop der Gesellschaft für Hochschulforschung (GfHf)

26. November 2012, Hannover Leibnizhaus

Am 26.11.2012 veranstaltete die Gesellschaft für Hochschulforschung (GfHf) in Kooperation mit dem HIS-Institut für Hochschulforschung (HIS-HF) einen Workshop zur OECD-Initiative AHELO (*Assessment of Higher Education Learning Outcomes*). Ziel des Workshops war es, Chancen und Risiken des Projekts aus wissenschaftlicher Perspektive zu diskutieren. An der Veranstaltung nahmen rund 50 Wissenschaftler(innen) und Expert(inn)en teil.

Nach der Begrüßung durch die Vorsitzende der GfHf, Prof. Dr. Margret Bülow-Schramm gaben drei Vorträge am Vormittag Inputs für die Arbeitsgruppen und Diskussionsrunden des Nachmittags. Elisabeth Hovdhaugen (NIFU, Oslo) berichtete von den norwegischen Erfahrungen mit AHELO. Norwegen gehört zu den 16 Ländern, die sich an der AHELO-Machbarkeitsstudie beteiligen und zwar in dem Teilbereich generische Kompetenzen. Die Ergebnisse der 2010 gestarteten Machbarkeitsstudie werden im Frühjahr 2013 auf einer Konferenz in Paris vorgestellt. Stefan Hornbostel (iFQ, Berlin) lotete in seinem Vortrag aus, unter welchen Bedingungen eine Teilnahme Deutschlands an AHELO, das gegenwärtig nicht an AHELO teilnimmt, aus der Perspektive der Wissenschafts- und Hochschulforschung sinnvoll erscheint. Hans Pechar (IFF, Wien) skizzierte abschließend das Spannungsfeld zwischen Wissenschaft und Wissenschaftspolitik, in dem sich Assessments wie AHELO bewegen (vgl. die Folien der Vorträge im Anhang).

Der Nachmittag diente dazu, das Thema der Kompetenzmessung im Hochschulbereich in drei Arbeitsgruppen diskursiv zu vertiefen. Die Gruppe um Elke Wild (Universität Bielefeld) erörterte die Frage, was die Hochschulforschung von der Schulforschung, in der internationale Vergleichsstudien zur Kompetenzmessung schon seit geraumer Zeit etabliert sind (z. B. PISA), lernen kann und wo die Grenzen der Übertragbarkeit liegen. Barbara Kehm (INCHER, Kassel) arbeitete mit der von ihr geleiteten Gruppe zur Frage der Leistungsfähigkeit von Rankings und Ratings. Dem internationalen Vergleich als Methode der Hochschulforschung widmete sich die dritte, von Dominic Orr und Kai Mühleck (beide HIS-HF, Hannover) moderierte Arbeitsgruppe. Die Ergebnisse der Arbeitsgruppen wurden anschließend im Plenum vorgestellt und diskutiert. Den Abschluss des Tages bildete eine Podiumsdiskussion, die zum einen die Ergebnisse des Tages aufgriff, zum anderen durch Statements der Podiumsteilnehmer(innen) noch einmal neue Aspekte in die Debatte einspeiste. Auf dem Podium diskutierten Sigrid Blömeke (HU Berlin), Holger Fischer (Universität Hamburg), Margret Schermutzki (Hochschule Aachen) und Frank Ziegele (CHE, Gütersloh). Die Vorträge können ebenso wie die Ergebnisse der Arbeitsgruppen und die Statements der Podiumsteilnehmer(innen) in der Dokumentation des Workshops detailliert nachgelesen werden, die im Frühjahr 2013 in der Reihe HIS:Forum Hochschule in Kooperation mit der GfHf erscheinen soll.

Als Ergebnis des eintägigen AHELO-Workshops und der dort geführten Diskussionen hält die Gesellschaft für Hochschulforschung (GfHf) als Fachgesellschaft der Hochschulforschenden im deutschsprachigen Raum Folgendes fest:

- Die GfHf hält eine umfassendere Erforschung des Kompetenzerwerbs und der Kompetenzentwicklung von Studierenden und Hochschulabsolvent(inn)en für dringend geboten. Die bislang auf diesem Feld vorherrschende Erhebung von Selbsteinschätzungen der Studieren-

den und Absolvent(inn)en ist durch Kompetenztests, wie sie in der Schulforschung bereits Standard sind, zu ergänzen. Die Gesellschaft macht hier ein wichtiges Forschungsdesiderat in der deutschen Hochschulforschung aus. Sie stellt aber zugleich fest, dass an die Messung von Kompetenzen im Hochschulbereich andere inhaltliche und methodische Anforderungen zu stellen sind als dies für den Schulbereich der Fall ist.

- Aus wissenschaftlicher Perspektive bieten Projekte wie AHELO die Chance, die Forschung auf dem Gebiet der Kompetenzmessung im internationalen Vergleich entscheidend voranzubringen. Ähnlich wie PISA für die Schulforschung könnte AHELO der Entwicklung geeigneter Messinstrumente für hochschulische Kompetenzen und damit der Hochschulforschung einen deutlichen Schub geben. Die Gesellschaft für Hochschulforschung erachtet es daher als wichtig, ihre wissenschaftliche Expertise – wo immer möglich – in derartige Projektkontexte einzubringen und sich an der fachwissenschaftlichen Diskussion aktiv zu beteiligen. Gerade der internationale Vergleich eröffnet – wie von AHELO angestrebt – die Chancen zum *peer learning* und zum besseren Verständnis der Besonderheiten der unterschiedlichen nationalen Kontexte.
- Die bislang über die AHELO-Machbarkeitsstudie und ihre Ergebnisse vorliegenden Informationen haben nach Ansicht der GfHf allerdings zugleich deutlich gemacht, dass eine Weiterentwicklung von AHELO geboten erscheint. Folgende Bedingungen müssten aus Sicht der Gesellschaft für Hochschulforschung erfüllt sein, damit eine deutsche Beteiligung an AHELO sinnvoll zur Weiterentwicklung der Qualität des Studiums beitragen kann:
- Prioritär erscheint der GfHf eine Klärung der Ziele: Welche Ziele sollen (und können realistischweise) mit AHELO erreicht werden? An welche Adressaten richten sich die mit AHELO erzielten Ergebnisse? Denkbar sind ganz unterschiedliche Ziele, die zugleich jeweils unterschiedliche Anforderungen an das Projektdesign und die Aufarbeitung der Ergebnisse stellen. So kann AHELO als Steuerungsinstrument in der Hochschulpolitik konzipiert werden oder der Information von Studieninteressierten über die Qualität der hochschulischen Ausbildung dienen. AHELO kann eingesetzt werden, um den Erfolg von Bildungsreformen zu messen, als Rückmeldesystem für Studierende über die eigene Leistungsfähigkeit oder Weiterbildungsbedarfe bei den Hochschullehrenden zu identifizieren. AHELO kann in erster Linie ein Projekt zur Weiterentwicklung der Kompetenzforschung sein, aber auch Marketinginstrument für die Hochschulen. Die Unterschiedlichkeit der möglichen Ziele zeigt deutlich, dass die (bislang nicht erfolgte) Zielklärung eine *conditio sine qua non* für eine erfolgreiche Weiterentwicklung von AHELO ist.
- Die kultursensible Messung von Kompetenzen in unterschiedlichen nationalen Kontexten wird von der GfHf als ambitioniertes wissenschaftliches Unterfangen erachtet, das weitreichender, komplexer und ressourcenintensiver Analysen bedarf. Die Gesellschaft hält hier weitere Grundlagenforschung für erforderlich. Die Methoden, die in der AHELO-Studie zur Kompetenzmessung über Ländergrenzen hinweg eingesetzt werden, sind gegenwärtig für Beobachter nicht hinreichend transparent. Dies wird sich erst mit der Veröffentlichung der Ergebnisse der Machbarkeitsstudie Mitte März 2013 in Paris ändern. Eine abschließende Bewertung der Leistungsfähigkeit von AHELO wird daher auch erst nach gründlicher Auswertung dieser Ergebnisse möglich sein. Soweit die Methodik jetzt schon bewertet werden kann, sieht die Gesellschaft insbesondere zwei Aspekte als kritikwürdig an:
- AHELO testet die Kompetenzen von Studierenden am Ende des Bachelorstudiums in drei Bereichen: fachübergreifende Kompetenzen (*generic skills strand*), disziplinspezifische Kompetenzen der Wirtschaftswissenschaften (*economics strand*), disziplinspezifische Kompetenzen

der Ingenieurwissenschaften (*engineering strand*). Es erscheint der GfHf als nicht zielführend, überfachliche Kompetenzen losgelöst vom fachlichen Bezugskontext zu messen. Die GfHf hält es insbesondere für die europäischen Länder – in denen die überfachlichen Kompetenzen von den Hochschulen innerhalb der fachwissenschaftlichen Ausbildung vermittelt werden – für notwendig, diese entsprechend auch im Testdesign an den disziplinären Kontext konzeptionell anzudocken.

- In der ursprünglichen Konzeption von AHELO war vorgesehen, nicht nur das zum Messzeitpunkt (Ende des Bachelorstudiums) erreichte Kompetenzniveau zu messen, sondern auch die Entwicklung der Kompetenzen zwischen einem Messzeitpunkt A (Eintritt in das Hochschulsystem) und einem Messzeitpunkt B (Verlassen des Hochschulsystems). Dieser sog. *value-added*, den man in diesem Kontext auch als Beitrag der hochschulischen Ausbildung zur Kompetenzentwicklung der Studierenden definieren könnte, stellt aus Sicht der GfHf den eigentlichen Mehrwert von AHELO im Vergleich zu anderen Studien dar. Von einem solchen Studiendesign sind wichtige Erkenntnisfortschritte in der Kompetenzforschung zu erwarten. Die Gesellschaft für Hochschulforschung plädiert daher dafür, diesen Ansatz bei einer Weiterentwicklung von AHELO wiederaufzunehmen, auch wenn sie sich der Tatsache bewusst ist, dass er ungleich komplexer ist als der gegenwärtig verfolgte.
- AHELO ist nicht als Ranking konzipiert, wird aber bereits jetzt von verschiedenen Akteuren als solches perzipiert. Die GfHf ist sich bewusst, dass auch die Politik Interesse an den Testergebnissen hat und an den Hochschulen eine curriculare Ausrichtung an den Testerfordernissen (*teaching to the test*) stattfinden könnte. Als Konsequenzen sind ein Eingriff in die Autonomie der Hochschulen einerseits, sowie eine Minimierung der Aussagekraft der Messergebnisse andererseits zu befürchten. Dieses Spannungsfeld ist nicht gänzlich zu lösen, da Indikatoren gestützte, international vergleichende Forschung immer als Ranking verwendet werden kann.

Für das Protokoll: *Margret Bülow-Schramm und Edith Braun*

Autor(inn)enverzeichnis

- Prof. Dr. Sigrid Blömeke** Professorin für Systematische Didaktik und Unterrichtsforschung an der Humboldt-Universität zu Berlin
sigrid.bloemeke@staff.hu-berlin.de
- Dr. habil. Edith Braun** Wissenschaftliche Leiterin des HIS-Instituts für Hochschulforschung, Hannover
braun@his.de
- Prof. Dr. Margret Bülow-Schramm** Professorin i. R. am Zentrum für Hochschul- und Weiterbildung (ZHW) der Universität Hamburg;
Vorsitzende der Gesellschaft für Hochschulforschung
buelow-schramm@uni-hamburg.de
- Dr. André Donk** Referent am HIS-Institut für Hochschulforschung, Hannover
donk@his.de
- Prof. Dr. Stefan Hornbostel** Leiter des Instituts für Forschungsinformation und Qualitätssicherung, Berlin;
Professor für Soziologie (Wissenschaftsforschung) am Institut für Sozialwissenschaften der Humboldt-Universität zu Berlin
hornbostel@forschungsinfo.de
- Dr. Elisabeth Hovdhaugen** Senior Researcher am Nordic Institute for Studies in Innovation, Research and Education (NIFU), Oslo
elisabeth.hovdhaugen@nifu.no
- Dr. Susanne In der Smitten** Wissenschaftliche Mitarbeiterin am HIS-Institut für Hochschulforschung, Hannover
smitten@his.de
- Prof. Dr. Barbara M. Kehm** Universitätsprofessorin für Hochschulforschung an der Universität Kassel (Internationales Zentrum für Hochschulforschung Kassel, INCHER-Kassel)
IBK.Kehm@t-online.de
- Tanja Meister** Referentin am HIS-Institut für Hochschulforschung, Hannover
meister@his.de

- Dr. Dominic Orr** Wissenschaftlicher Mitarbeiter am HIS-Institut für
Hochschulforschung, Hannover
orr@his.de
- Prof. Dr. Hans Pechar** Leiter der Abteilung Hochschulforschung an der Fakultät für Interdisziplinäre Forschung und Fortbildung (iff)
Universität Klagenfurt, Standort Wien
hans.pechar@uni-klu.ac.at
- Margarete Schermutzki** Higher Education Expert Tuning
Tuning Academy, Bilbao/Groningen
mschermutzki@yahoo.com
- Prof. Dr. Robert Wagenaar** Director of Undergraduate and Graduate Studies an
der Universität Groningen
r.wagenaar@rug.nl
- Prof. Dr. Frank Ziegele** Geschäftsführer des CHE Gemeinnütziges Centrum für
Hochschulentwicklung GmbH;
Professor für Hochschul- und Wissenschaftsmanage-
ment an der Fachhochschule Osnabrück
Frank.Ziegele@CHE.de

HIS, Goseriede 9, 30159 Hannover
Postvertriebsstück, Deutsche Post AG, Entgelt bezahlt, 61246

Herausgeber dieser Ausgabe:

Gesellschaft für Hochschulforschung (GfHf),
HIS Hochschul-Informationen-System GmbH

Herausgeber der Reihe HIS:Forum Hochschule:

HIS Hochschul-Informationen-System GmbH
Goseriede 9 | 30159 Hannover | www.his.de

Postfach 2920 | 30029 Hannover
Tel.: +49(0)511 1220 0 | Fax: +49(0)511 1220 250

Geschäftsführer:

Dipl.-Phys. Wolfgang Körner

Vorsitzender des Aufsichtsrats:

Prof. Dr. Andreas Geiger

Registergericht:

Amtsgericht Hannover | HRB 6489

Umsatzsteuer-Identifikationsnummer:

DE115665155

Verantwortlich:

Dipl.-Phys. Wolfgang Körner

Erscheinungsweise:

In der Regel mehrmals im Quartal

Hinweis gemäß § 33 Datenschutzgesetz (BDSG):

Die für den Versand erforderlichen Daten (Name, Anschrift) werden elektronisch gespeichert

ISBN 978-3-86426-024-7

