JUDITH GULIKERS AND MARTIN MULDER

CONFERENCE "MODELLING AND MEASURING COMPETENCIES" BERLIN, 24-25 FEBRUARY, 2011.

In February, 2011 I visited Berlin to attend the special conference on "Modelling and Measuring Competencies in Higher Education" organised by the Humboldt University of Berlin and the Johannes Gutenberg University of Mainz. It was a special conference as it only addressed this topic for two days and had mainly invited (keynote) speakers. While the audience consisted of mainly German participants, leading researchers from universities and testing institutes in mainly Germany, USA, Australia presented their views and work. After three elaborate keynotes in the morning, the afternoon of both days consisted of a panel discussions or a, so-called, town-hall meeting in which 3 to 5 researchers introduced their work in 15 minutes followed by a lively and interactive discussion. During lunchtime on Friday there was a poster round presenting 14 posters coming mainly from Germany and some from Finland. It was a very interesting, inspiring, but also confronting experience. I would like to discuss two crucial controversies that were highly illuminated at these two day meeting where the first day mainly focused on large-scale, high stakes assessment versus a second day that paid much more attention to the individual student in the assessment process:
1. measuring the cognitive aspects of competencies or something more?
2. curriculum validity versus professional/labour market validity
   These discussions, and the positions researchers take in this respect or the frame of reference they are in (either large-scale (high stakes), or the individual student level), has, to my opinion, a big impact on both the modelling and measurement of competencies (or competences) as well as on the impact such an assessment can or should have on the teaching and learning process.

THE FIRST DAY: LARGE-SCALE (HIGH STAKES) MEASUREMENTS AT THE INSTITUTIONAL LEVEL

The openings words of Prof. Sigrid Blömeke informed the audience about a large German Ministry of Education and Research funding initiative called "Modeling and Measurement of Competencies in Higher Education". This initiative stimulates new, creative but also fundamental research emphasising *more evidence-based innovation* in (Higher) Education teaching and learning. 94 proposals from various disciplines were submitted, showing the interest and relevance of this topic and the necessity of doing research in this field.

After the opening word, the first day was filled with contributions from mainly German and North American researchers, except for the keynote of Karine Trembley, senior survey Manager of the "Assessment for Higher Education Learning Outcomes (AHELO)" international project of the OECD. Overall, this day was characterised by viewing assessment as a large-scale, high stakes undertaking, mainly purposing comparing (or even ranking) institutions at a national or international level. Probably because of this purpose, all assessments were written tests, containing often even multiple choice formats, mainly gauging the cognitive aspects of competence. This issue was heavily discussed in the panel meeting in the afternoon, discussion four German projects. The majority of these German research projects defined competence by its 'narrow definition' described by Klieme and Leutner (2006): "context specific cognitive dispositions that are acquired and needed to successfully cope with certain situations or tasks in a specific domain". Blömeke adds to these cognitive aspects the importance of taking emotional / motivational aspects into account and therefore assesses competence not only by addressing students' cognition, but also their beliefs. The choice to only deal with the cognitive aspects was defended by arguing that these elements are objectively measurable through written items, which is almost inevitable in large-scale assessment used to compare institutions. However, an additional argument was made for 'curriculum validity': measuring these elements that are also present in the curriculum. The assessments were drawn up after consultation with faculty members from various HE institutions discussing the contents of their curriculum. Involvement of the labour market was not an issue, as the labour market for the HE graduates was argued to be too vague and too broad. The question immediately popping up in my mind was obviously:

> If the purpose is to innovate Higher Education (see the funding initiative), than assessing only these elements that are present in the current curriculum will not stimulate innovating the curriculum, will it?

Prof. Richard Shavelson, a well-known American assessment specialist, might offer us a way out of assessing only cognitive aspects in large scale tests: he offered us a technical inside into assessing competencies through written test items, that are also authentic performance assessment tasks representative of competences used in the real world. He argued that the starting point of developing such an assessment should be a careful description of the criterion domain of behaviour in

2

real life (see also Shavelson, 2010) and thus not the current curriculum of the HE institutions. His technical presentation was made more concrete by Roger Benjamin at the second day of the conference, who discussed the Collegiate Learning Assessment (CLA) used all over USA higher education institutions (see also Benjamin, Chun, & Shavelson, 2009). This is a large-scale high stakes assessment of critical analysing and evaluation, problem solving, writing persuasion and writing mechanics through an array of written, but real life, performance tasks scored by trained computer systems. The degree to which written tasks can be called performance tasks can be debated, but he got me convinced at this time that these are definitely more performance, real-life and competence oriented for the HE context, than the impression I got from the examples I heard on the first day of this conference. Next to the focus on real life performance tasks of generic competences instead of disciplinary content, the CLA also stresses the necessity of holistic, instead of analytic or atomized, scoring. For each performance tasks, three integrative (or holistic) model answers are developed characterising high, moderate or low performance. Maybe even more important, from my point of view, was Benjamin's emphasis on using the CLA as an improvement instrument for the teaching and learning process in the HE institution. The CLA feeds back at the institutional level, saying something about the value added of a certain institution, and thus its curriculum or educational program. In this way, the CLA does not aim at curriculum validity (that is, assessing only what it taught in the curriculum), but it aims at offering institutions handles to improve their educational program to address real life challenges their graduates will face in their future. During the discussion, Blömicke addresses the issue of taking students' entry level of competence into account to say something about the value added of the HE institutions. This issue will be further discussed on day two by Prof. Spiel.

THE SECOND DAY: THE INDIVIDUAL STUDENT AND COMPETENCE ASSESSMENT

The above discussion focused mainly on large-scale, written assessments, with the aim of comparing, ranking and improving at the institutional level. Various speakers at the second day of the conference, however, strongly emphasised the need to look at the individual (student) level when talking about modelling and measuring competences. Professor Michaela Pfadenhauer (Germany) discusses the word competence from a sociological perspective and builds up an argument for what this means for the modelling and measurement of competences. She advocates a wide definition of competence that includes the social aspects of competence, which is inherently connected to the individual, his interpretation of the world and his feeling of responsibility for acting. Being competent, in her words, means that the actor combines "können, wollen, und dürfen" into repeated and responsible action in various situations. This entails that modelling and measuring competence cannot be done without carefully examining actors' actions and including the actors' own perspective and reflection on his performance. If we

follow this line of reasoning, she argues for the impossibility of objectifying competence and using standardized measurements and assessments that do not involve a dialogue with the actor and examining (through introspection of retrospection) his/her actions. She takes this even further in arguing that the actor should have the decisive vote in determining whether or not he/she possesses a certain competence or not. An obvious point of discussion is raised, namely various scientific studies showing the lack of reliability of self-assessments. Pfadenhauer responds to this issue by saying:

> I am not arguing that you are on the safe side of competence assessment when you use self-assessments. I argue that we should reframe how we view and use self-assessments. The self-assessments that are often used in practice and research are not addressing competence as I see it, namely as a personal and responsible implementation of action. If you see competence like this, than self-assessment through introspection based on various performances is inevitable.

The problem, she states, is that this perspective on competence development and measurement requires (German) both education and the accountability system to change, which are both not happening. This is a problem recognised by other researchers and in various other counties as well (e.g., Knight (UK) and Kvale (Denmark), 2007) and it is safe to say that these are problems faced in the Dutch movement towards competence based assessment and education as well.

Professor Sadler (Australia) agrees with Pfadenhauer on the importance of involving students in the assessment and development of their competences and having dialogues with students on the meaning of competence and being competent within a certain area or task. He also strongly stresses the need for an holistic approach to assessment of competence (see also Sadler, 2009):

> Decomposing competence into manageable (or even atomized) components in order to facilitate judgements may have some interim value in certain contexts, but the act of decomposition can obscure hoe a practitioner would work the various bits together to form a coherent whole".

In this holistic approach and the need for assessments of competence to relate to the professional world he agrees with the arguments of Shavelson and Benjamin, of the CLA. However, contrary to Pfadenhauer who places the main responsibility for competence assessment with the actor and the CLA that uses written performance tasks scored by computer, he elaborates on the crucial role of complex performance tasks that can only by judged by knowledgeable judges who have calibrated their views on what competence means. This opens up an new discussion on the pros and cons as well as the (im)possibilities of human judges in complex, open ended assessment tasks. The link to assessment quality, increased focus on the assessment process and procedure as well as the professionalism of human judges (i.e, teachers) is made.

Professor Christiane Spiel (Austria; second day of the conference) flies in from a bit different viewpoint, namely that of program evaluation and evaluation

research. Her talk combines the institutional perspective as well as the student perspective of competence assessment discussed above, by discussing the issue of evaluating the value added of a HE curriculum for the development of competences of students. She emphasised using outcome evaluations that evaluate the extent to which programs achieve their goals, which seems to be obvious, but she stresses that goals should address both generic and domain-specific competences instead of only disciplinary knowledge. To examine the value added, evaluation systems should compare these competences of graduates and freshmen (baseline data) as well as compare graduates competences with the defined graduates profiles.

Professor van der Velden (the Netherlands) makes this value added discussion more concrete by discussing, again large-scale but this time not high stakes, graduate surveys. He argues that these surveys should measure both cognitive and non-cognitive aspects of students' competences as developed through a certain curriculum containing certain characteristics and relate these to the outcomes students achieve in the labour market. Then, these data can provide useful information on (1) what type of outputs (i.e. competencies? Or something else?) a HE institution should be focussing on based on requirements of the dynamic world of work, and (2) what HE institutions can do to better foster these outputs. Results show that the HE labour market strongly emphasises professional expertise including both domain specific as well as generic academic competences of which this last category increases in importance to equip graduates to deal with uncertainty in society and become functionally flexible in the labour market. While van der Velden bases his arguments on large scale written survey data combining and comparing HE institutions, he stresses that at the institutional level, the development and measurement of competence requires other assessment methods than written, specifically multiple-choice, tests as these will not stimulate students to develop of generic academic competences, nor will these tests stimulate teachers to educate for generic competence development. This immediately shows, what van der Velden calls, the assessment gap that many HE institutions are in: most HE institution still mainly rely on MC-testing of disciplinary skills instead of (also) using other types of assessments and assessing professional expertise in terms of both domain-specific and generic academic competencies representative of the labour market.

Thus, there is a lot to be done to improve the modelling and measurement of competences in HE that drive institutions to innovate towards professional validity. This conference showed that various countries have different perspectives on competences (or competencies), their modelling and the qualities and form of their measurement. This conference showed that there is a large scale side to competence assessments; a cognitive, a performance and a social or beliefs side; an individual student and an institutional side; a curriculum and a labour market side; a dialectical/reflective versus a standardized/one-size-fits-all side; and moreover, a summative, comparing and ranking purpose and a formative purpose feeding back to the institution, program or student level stimulating improvement and development. However, independent of your viewpoint or purpose, all participants

agreed that competence assessment is increasing in its relevance and importance both at national and international level, shown in the big funding initiative of the hosting countries' Ministry of Education. However, it is fraught with controversies and difficulties as it challenges how we define quality of (higher) education, how (higher) education institutions change, or should change, to address these challenges, and how institutional evaluation systems (like Spiel was talking about), as well as external accountability bodies examine, grant, and stimulate improvement of the quality of the educational curricula, teaching and assessments.

REFERENCES:

Benjamin, R., Chun, M., & Shavelson, R. (2009). *Holistic tests in a sub-score world: the diagnostic logic of the collegiate learning assessment (CLA)*. Retrieved at February 28th, 2011, from http://www.collegiatelearningassessment.org/

Knight, P. (2007). Grading, classifying and future learning. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the long term*. New York: Routhledge.

Kvale, S. (2007). Contradictions of assessment for learning in institutions of higher learning. In D. Boud & N. Falchikov (Eds.), *Rethinking assesment in higher education: Learning for the long term.* New York: Routhledge.

Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 45-63): Springer-Verlag.

Shavelson, R. (2010). On the measurement of competency. *Empirical research in vocational education and training, 2*(1), 41-64.

Judith Gulikers and Martin Mulder
Chair group Education and Competence Studies
Wageningen University, The Netherlands