

To be published in *Assessment & Evaluation in Higher Education*

DOI: 10.1080/02602938.2012.755150

Evaluating the level of degree programmes in higher education: conceptual design

Trudy Rexwinkel*, Jacques Haenen and Albert Pilot

Faculty of Social and Behavioural Sciences, Department of Education, Utrecht University, Utrecht, The Netherlands

Abstract

Evaluating the level of degree programmes became crucial with the Bologna Agreement in 1999 when European ministers agreed to implement common bachelor and master degree programmes and a common system of quality assurance. The European Quality Assurance system demands evaluation of the degree programme level based on valid and reliable outcomes. To meet this standard, an *Educational Level Evaluator* had been conceptually designed within the framework of the Design Research Methodology: the Research Clarification identifies the problems that need to be solved and the criteria against which the outcome of the research is evaluated. The Descriptive Stage focuses on acquiring understanding of the instrumentation's key by conceptualising the degree programme level and developing four critical factors. The level concept has to be clear and the instrumentation should be reasoned transparently, empirically proven, and feasible. In the Prescriptive Stage the critical factors are used to address the development of the conceptual design of the *Educational Level Evaluator*, which is basically a procedure aiming to evaluate the degree programme level validly and reliably. It is concluded that the design provides measures and actions leading to validity and reliability, and that it now needs to be empirically proven. For this reason, empirical studies are undertaken.

Keywords: degree programme level, construct validity, conceptual design, design research methodology

Introduction: Research Clarification

Evaluating the level of degree programmes became crucial with the Bologna Agreement in 1999 when European Ministers of Education agreed to implement the requirements of the common bachelor and master structure, and the common system of quality assurance: accreditation. The main characteristics of accreditation are that it is based on the degree programme level and that it is a condition for finance. For degree programmes it is necessary to prove the real level with quantitative outcomes. This appears to be a complicated assignment for degree programmes. These criteria were applied by the [European Association for Quality Assurance \(ENQA\)](#), which was mandated by the European Ministers of Education to implement European Quality Assurance in order to realise the system of accreditation.

The main challenge of this study is to design an *Educational Level Evaluator* (ELE), a procedure with instrumentation to evaluate the degree programme level validly and reliably. The *ELE* is designed within the framework of the Design Research Methodology (DRM) ([Blessing & Chakrabarti, 2009](#)), which is explained in the Method section (Figure 1). In the first DRM stage, the Research Clarification, the problem that needs to be solved is introduced, its academic relevance is indicated, and the aim of this study is made clear. The Research Clarification is completed with research questions against which the outcome of the design is evaluated.

Academic relevance

The aim of this study is to advance our knowledge about the degree programme level and how to evaluate it. Instrumentation to evaluate the degree programme level validly and reliably is not yet available.

Initiatives have been undertaken and instruments developed, but these have found no commitment and were not implemented. For example, the Quality Assurance Agency for Higher Education (QAA) in the UK publishes subject benchmark statements, indicating the degree level that higher education institutions are expected to consult when designing, delivering and reviewing programmes. However, within the context of higher education, this is considered by some to be unacceptably bureaucratic ([Bellingham 2008, p. 265](#)). Despite the large supply of benchmark

statements, systematic evaluation of the effectiveness of subject benchmark statements in assuring academic quality and standards is lacking (p. 274).

European policy makers developed the Dublin Descriptors to describe the levels of bachelor and master degrees ([Bologna Working Group, 2005](#)). It was intended that these qualifications should apply in higher education in the European countries to enable cross-national comparison of the degree programme level. International policy makers agreed with the generally formulated descriptors. However, these are worded in broad terms, allowing too many interpretations, which lead to ambiguity and hinder the cross-national comparison of the degree programme level.

The Organisation for Economic Cooperation and Development (OECD) had explored feasibility studies to determine the extent to which HEIs effectively prepare their students for participation in increasingly knowledge-based economies. The OECD Assessment of Higher Education Learning Outcomes project concluded that ‘in most countries, assessment results are inaccessible and [...] if HEIs would specify the expected student outcomes explicitly and in a measurable way, comparative assessment of learning outcomes would become feasible’ ([Nusche, 2008, pp. 5-9](#)).

The problem of the degree programme level is complicated ([Cosgrove, 2011](#)) and “remains elusive” ([Ewell, 2010, p.173](#)). The evaluations of the Bologna Process ([Westerheijden et al., 2009](#)) make it obvious that the aim of the demand for more objective proof of realising the degree programme level is more complicated and will take longer than a decade. The relevance and urgency of these evaluations are recognised by the European Ministers to a certain extent and they postponed the deadline from 2010 to 2012 for achieving these Bologna aims.

Problems experienced by accreditation organisations are related to the core assignment, i.e., evaluating. To evaluate the degree programme level “requires in-depth subject knowledge, international orientation and evaluation experience [...]”. Considering the quality of evaluators of higher education [...] there might be room for the introduction of standards for evaluator selection.” ([Harris-Huermert, 2008, p. 55](#)). “Expertise about evaluation is not present in accreditation committees.” ([Sandahl et al., 2006, p. 9](#)). “Subjective assessments of programme content are observed.” ENQA requires systematic assessment of student learning and use of empirical evidence to guide their efforts to improve quality. “Without this type of concrete evidence [...]

alternative efforts to improve academic quality are likely to be wasteful and ineffective” (Konrad et al., 2007, p. 23). The aim of this study is to design an instrumentation for systematic evaluation of the degree programme level, with the following research question: *What instrumentation can be designed to evaluate the degree programme level validly and reliably?* This general research question is detailed with four sub-questions:

1. How can the instrumentation be designed?
2. What is the concept of the degree programme level?
3. What are the critical factors of the instrumentation?
4. What does the instrumentation look like?

The research questions refer to the conceptual design of the instrumentation, which aims to evaluate the degree programme level validly and reliably. With this aim, this study is organised within the framework of the Design Research Methodology (DRM), which begins with the Research Clarification (covered by this Introduction section) leading to the above-mentioned research questions. Then the DRM framework is explained under Method. Apart from the Research Clarification, the DRM framework consists of a Descriptive Stage which focuses on the key of the instrumentation evaluating the degree programme level and outlines the critical factors of the instrumentation. In the Prescriptive Stage, the critical factors address the development of the conceptual design. This study closes with conclusions and discussion.

Method: Design Research Methodology

Within the framework of the Design Research Methodology (DRM; Blessing and Chakrabarti (2009), the conceptual design of the *Educational Level Evaluator* is developed. The process of designing implies the application of certain principles, terminology and methodology. Blessing and Chakrabarti ascertained that designing is multidisciplinary but is characterised by a lack of common terminology, benchmarked research methods, and common research methodology. In DRM they described the common multidisciplinary stages by designing activities. As these descriptions

support this study's target, DRM is also used for designing conceptually the Educational Level Evaluator.

The Design Research Methodology consists of four stages and has a cyclical nature. The first cycle contains the stages: Research Clarification, Descriptive Stage, and Prescriptive Stage. The Descriptive Stage II is described after analysis of the empirical data.

Insert Figure 1 about here

In the Research Clarification the approach and methods are introduced to support the initial stage of the design project. The aims are to identify and refine the research problem that is both academically and practically worthwhile and realistic. The Research Clarification is completed with research questions against which the outcome of the design can be evaluated.

Descriptive Stage I aims to recognise and clarify in more detail the issues of the existing situation to address the development of the desired situation. The Descriptive Stage builds on the deliverables of the Research Clarification and provides a sound base for the Prescriptive Stage. This requires a deep understanding of the key of the instrumentation. The main notion of the instrumentation, i.e. the degree programme level, is conceptualised clearly in the Descriptive Stage, and the critical factors, i.e. causes with an essential influence on the instrumentation's performance, are identified.

The Prescriptive Stage focuses on how, starting with the results from the Descriptive Stage, to proceed for the development of the design procedure with instrumentation to deal with the influence of (some of) the critical factors found in the Descriptive Stage. The conceptual design of the *ELE* is developed by addressing the critical factors in a systematic way and realising this to such a level of detail that it can be evaluated.

In Descriptive Stage II empirical studies can be undertaken to establish *ELE* meets the criteria of validity and reliability.

Descriptive Stage

Aim of the Descriptive Stage is to outline and specify the concept of the degree programme level and the critical factors of the instrumentation.

Conceptualising the degree programme level

One of the most important aims of degree programmes in higher education is that students learn to think on complex disciplinary problems at the bachelor and master level. Higher education is the main place for students to develop the thinking power that is considered as the core of the degree programme level. In this study this level is conceptualised with disciplinary thinking and professional attitude in this study. Both components are related to the domain and profession the study is addressing.

Disciplinary thinking refers to higher order thinking processes (analysing, evaluating, creating) on complex disciplinary problems. Bloom identified such thinking processes on the higher scale of his taxonomy of educational objectives ([Kek & Huijser, 2011](#); [Robinson, 2011](#); [Bransford, Brown & Cocking, 2000, p. 31](#)).

Disciplinary thinking applies also to ‘conceptual understanding’ ([Hattie \(2009, p. 248\)](#)). [Sternberg \(2003\)](#) emphasised that students should not only be educated as content experts with a large knowledge base, but as experts who can think analytically, creatively, and practically within their discipline. They need creative thinking to generate ideas, analytical thinking to evaluate those ideas, and practical thinking to implement the ideas and convince others of their value. It is a great challenge to design such courses, as is clear in, e.g., business research ([Chen, 2010](#)) and psychology ([Alibali, Spencer, Knox & Kita, 2010](#)).

Professional attitude is a component of the degree programme level too. This attitude refers to higher order processes of organising and characterising attitudinal aspects of the profession or field the study is addressing, e.g., the accuracy of the bookkeeper, the hospitality of the hotel manager, or the independence of the scientist. This could be exemplified with the following description of a bachelor psychologist: “A bachelor psychologist is able to handle ethical dilemmas in the professional researcher’s field. He is able to handle these rationally and emotionally in correspondence with the professional code of psychologists” ([American](#)

[Psychological Association, 2012](#)). The professional attitude is elaborated further in the conceptual design.

In summary, the degree programme level is conceptualised as having two components: a cognitive ('disciplinary thinking') and an attitudinal ('professional attitude') component.

Critical factors of the instrumentation

Critical factors have an essential influence on the instrumentation's performance. The European Quality Assurance System requires the degree programme level to be demonstrated validly and reliably. This appears to be a complicated problem for the degree programmes as they need accreditation for finance, while the instrumentation to evaluate the level of degree programmes validly and reliably is not yet available ([Bellingham, 2008](#); [Westerheijden et al., 2009](#)). Based on the indications from the Research Clarification, four critical factors are distinguished about the clarity of the concept (1) and the instrumentation (2), and about the instrumentation's empirical evidence (3) and feasibility (4).

The concept of the degree programme level has to be clear. The level concept is a central part of the intended instrumentation and is developed from a clear vision of higher education. The conceptual content cannot be limited by general terms ([Bologna Working Group, 2005](#)). The conceptual content is derived from educational research (e.g., [Cosgrove, 2011](#)) and the concept should be suitable for providing evidence of student learning ([Ewell, 2010](#)).

The instrumentation has to be reasoned clearly. The arguments for the instrumentation are derived from conceptual content more than from empiricism, analyst's techniques or instrumental claims. Furthermore, the arguments focus on the coherence between the constituents.

The instrumentation has to be empirically proven. The *ELE* provides concrete empirical evidence of validity and reliability from the degree programme level ([ENQA 2009](#); [Westerheijden et al., 2009](#)). The instrumentation's design has to be scientifically sound and resistant to the many critical stakeholders' eyes. In a previous study it was demonstrated that grades are suitable indicators of the degree

programme level. In this study a second track is examined with a professional questionnaire.

The instrumentation has to be feasible. The feasibility of the instrumentation is principally determined by the quality of the system of accreditation. The system of accreditation has not existed for long. Consequently, the system is not yet perfect. As accreditation is essential for degree programmes, the system has to be professionalised in order that the *ELE* can be feasible. Professionalising requires a systematic approach, transparency about the criteria used, skilled evaluators with in-depth subject knowledge, international orientation, evaluation experience (Harris-Huermann, 2008), and expertise in evaluation (Sandahl et al., 2006; Konrad et al., 2007). The instrumentation is also feasible if it is not ‘unacceptably bureaucratic’ (Bellingham, 2008) for the degree programmes. The deliverables of the Descriptive Stage address the development of the conceptual design.

Prescriptive Stage: Conceptual design

After the conceptualisation of the degree programme level, the design of the *Educational Level Evaluator* is continued by operationalising, and measuring this level.

Operationalising the degree programme level

In correspondence with the second critical factor, the degree programme level is operationalised clearly for degree programmes per domain. The instrumentation’s design is content-driven. First, themes are created consistent with disciplinary thinking and professional attitude. The themes are developed after thorough consideration as they are the addressing element for evaluation of the degree programme level. The themes need to be created systematically, in order for them to generate evidence of the measurement’s validity (Field, 2009, p. 783). This addressing element is indicated in Figure 2.

Insert Figure 2 about here

Operationalising the degree programme level concept requires expertise of the language, knowledge and culture of the degree programme(s), implying that *ELE* is carried out in collaboration with degree programmes, i.e., disciplinary teachers and expert members.

To create the themes a specific method was developed: for objectivity, it is analysis-based. The existing themes in the current curriculum, the external surroundings, the professional and the academic fields are analysed with the question: What findings and developments are important for the degree programme level?

The developed themes refer to findings and developments. Each theme is underpinned with peer-reviewed literature. The expert members discuss the themes and their result is presented for face validity, i.e., various stakeholders of the degree programme(s) such as the dean, teachers and students, as a serious lack of credibility can influence the participants' efforts in the performance (Kane, 2006, p. 36).

The validated themes are specified with intended learning outcomes that are statements, written from the student's perspective, indicating the degree programme level (Biggs & Tang, 2007, p. 55). It is essential to express the appropriate degree of complexity in the formulations of the learning outcomes. The SOLO taxonomy is applied for the degree of complexity in learning outcomes regarding 'disciplinary thinking'. The Attitude Development Scheme is used for the degree of complexity in learning outcomes of 'professional attitude'.

SOLO stands for Structure of the Observed Learning Outcome and provides a systematic way of describing how a learner's performance grows in complexity when mastering many academic tasks. Biggs and Collis (1982) developed the taxonomy with five levels. The taxonomy was applied to various disciplines with the aim of investigating how students learn to master their academic tasks.

Consistency is the central notion in discriminating the levels of this taxonomy. This evolves from no consistency in the student's thinking, via no inconsistency, into resolving inconsistencies. The taxonomy's levels are explained from the fifth into the first level in Table 1. At the taxonomy's fifth level, the student is able to resolve inconsistencies. He feels no need to give closed decisions, holds open conclusions, and is qualified to allow logically possible alternatives (deduction and induction). He can generalise to situations not experienced. The student on the fourth level shows

consistency within the given system, but since his closure is unique, inconsistencies may occur when going outside the system. He can generalise within a given or experienced context using related aspects (induction). Although the student at the third level has a feeling for consistency he can be inconsistent because he closes too soon on the basis of isolated fixations on data, and so can come to different conclusions with the same data. At the second level, the student jumps to conclusions on one aspect and so can be very inconsistent. At the first, pre-structural level, the student feels no need for consistency and closes without even seeing the problem. He is confused, denies the problem or uses tautology (pp. 24-25). [Biggs and Tang \(2007\)](#) elaborated the taxonomy for learning at university and presented verbs for intended learning outcomes from the SOLO taxonomy.

Insert Table 1 about here

For the development of the Attitude Development Scheme, the SOLO taxonomy was a source of inspiration. The scheme's content was derived from the taxonomy of [Kratwohl, Bloom, and Masia \(1974\)](#) and the self-regulation cycles described by [Zimmerman \(2006\)](#). The degree of complexity in the attitudinal development has been tested for inter-rater reliability. The Attitude Development Scheme supports the wording of the intended learning outcomes that express the attitude of the profession or field the study is addressing.

The Attitude Development Scheme is reflected in Table 2 and provides an approach that indicates how a learner's performance develops in professionalism when characterising the attitudinal aspects of the profession or field. This development runs through five levels. At the fifth level, the student 'characterises' the various aspects of professional attitude, meaning that the aspects are consistently placed into a hierarchical control of his own attitude. In the fourth cycle, the student 'organises' more than one aspect into a system without inconsistencies within the given system. With 'valuing', the student demonstrates that the required aspects of the professional attitude gradually become observable and are developing to consistency on the third level after the student gradually internalises, inconsistently, characteristics of the specific attitude. On the second level, he 'responds' by tending

toward the attitude's aspect as intended, but is very inconsistent. On the first level, he becomes aware of the aspects of the professional attitude when he 'receives' descriptions of the professional attitude he is being educated for.

For example, professional attitude of the nurse appears in 'discretion'. Discretion is an umbrella term referring to respect, accuracy, and skilful, correct, appropriate, sensible, reserved consideration for the patient. These attitudinal aspects are expressed in the various actions of a nurse: when nursing the patient, and in communication with the patient and others (Rexwinkel, Haenen & Pilot 2012b).

Insert Table 2 about here

After the themes are specified into learning outcomes, they are validated by other expert members of the degree programme(s), who answer the question to what extent the learning outcomes are a good representation of the degree programme level that is measured with interrater reliabilities (Cohen's Kappa coefficient) (Cheniaux, Landeira-Fernandez, Versiani 2009). The number of learning outcomes is indicated by Cichetti's (1976) rule of $2n^2$. This rule states that the number of learning outcomes in a category needed for a reliable interpretation of a computed Kappa should be $2n^2$ or more. For a reliable test, in the case of four categories a minimum of ($2 \times 4 = 8 \times 2 =$) 16 items are necessary.

Measuring the degree programme level

Corresponding to the third critical factor, the instrumentation has to be empirically proven. Apart from various stakeholders (face validity) and expert members (interrater reliability K), validity is needed also from recent graduates who have achieved the degree programme level and from students who are still acquiring it. A questionnaire is constructed in order to measure the degree level. To optimise its validity, various actions are taken. The intended learning outcomes are transformed into items (Smith, 2003, p. 88) and a scale is determined. The numbers on the scale, mostly five-point, have normative meanings and are defined in terms with which students are familiar. The argument is that such a scale gives the user immediate normative information (Angoff, 1971, p. 528).

The questionnaire is pre-tested on students. The candidates should understand immediately the intended core of the question. If they hesitate, then there may be ambiguities. Difficult syntax, mismatching questions and answers, overlapping categories, missing categories and other problems can be traced with the coding systems of [Snijkers \(2002, p. 71\)](#).

The data collection requires careful organisation handled by administrators and graduates' coordinators who are aware of validity. Some major points in this process are: The respondents' anonymity needs to be guaranteed. The relationship between the degree programme and the graduates is maintained well. It might be necessary to develop manuals and to communicate via webinars to ask questions, or assess other aspects of the organisation.

Before analysis, the dataset is *inspected* for aspects of validity to determine that it is suitable for analysis, e.g., the size of the dataset, the scale, the dispersion of the data, and missing values.

The qualitative and quantitative *representativeness* are established. The (absolute) sample size is important, as well as the background information on the respondents ([Rutowski and Rutowski, 2010](#)).

Construct validity is crucial as it indicates to what extent the targeted construct has been measured. The construct cannot be measured directly. Constructs are not observable, but hidden behind the variables that are measured and discovered by construct analysis ([Harkness, Van de Vijver & Johnson, 2003](#)). This is a type of factor analysis suitable for studies of data, assessments, and evaluations of treatments ([Cronbach, 1992, p. 391](#)). The method for delivering evidence for construct validity is Principal Component Analysis with Varimax rotation. The data should meet the norm for construct validity, i.e., the conceptualised themes are measured as existing constructs ([Creswell, 2007, p. 204](#)) with components eigenvalues ≥ 1 explaining $\geq 50\%$ variance indicating what is really is measured.

If the data meet the norm for construct validity, *scale reliability* is analysed with alpha coefficients. These should measure ≥ 0.60 , the norm for measurement in groups ([COTAN, 2011](#)).

If the measuring instrument meets the norms of construct validity and scale reliability, the outcomes are verified on *concurrent validity*. This implies the

demonstration of the correlation of these outcomes with the outcomes of another instrument that is validated, measures similar constructs and is used in the same period of time.

If the outcomes are valid and reliable, z-scores are calculated from the measured themes and the degree programme level can be evaluated. It is established what themes of the study programme meet the norm and explanations for the outcomes are considered. The results may please or concern the degree's members. The seriousness of possible problems is determined and measures for improvement are developed and planned.

Conclusion and discussion

This article describes the development of the conceptually designed *Educational Level Evaluator (ELE)*, a procedure with instrumentation aiming to evaluate the degree programme validly and reliably. The process of designing and the design itself were examined with four questions:

How can the instrumentation be designed? The aim of this study was to advance our knowledge about the degree programme level and how to evaluate it. Meeting this aim, the instrumentation was conceptually designed using the Design Research Methodology (DRM) framework, including phases and actions described with multidisciplinary principles, terminology, and methodology. The Research Clarification identified the problems that need to be solved and the research questions against which the outcome of the research was evaluated. The Descriptive Stage focused on acquiring understanding of the instrumentation's essence: the degree programme level and the critical factors. In the Prescriptive Stage the critical factors addressed the development of the conceptually designed *ELE*.

What is the concept of the degree programme level? The degree programme level was conceptualised as having a cognitive and an attitudinal component. The cognitive component refers to disciplinary thinking, implying higher order thinking (analysing, evaluating, creating) on complex disciplinary problems. The attitudinal component refers to processing (organising, characterising) the main characteristics of the profession's attitude or the field the study is addressing, e.g., the integrity of the accountant.

What are the critical factors for the instrumentation? The critical factors for the instrumentation are contributory causes with the strongest influence on the instrumentation's success. Four critical factors were identified: the level concept as well as the instrumentation has to be content-driven, the instrumentation should moreover be empirically proven, and feasible.

What does the instrumentation look like? The *ELE* is a procedure with instrumentation aiming to evaluate the degree programme level validly and reliably. It is a generic model that can be elaborated by degree programmes per domain. It provides the steps of conceptualising, operationalising, and measuring the degree programme level with various supporting instruments, such as the SOLO taxonomy and the Attitude Development Scheme. The *ELE* provides face validity from various stakeholders, content validity by expert members and construct validity from graduates and students.

The purpose of this study is to contribute to the resolution of the complicated but urgent problem of the degree programme level and its evaluation. A vulnerable point might be the number of responding graduates in the data collection. Generally graduates are willing to participate, especially if the relationship between them and the degree programme has been maintained well.

The *ELE* was evaluated using the research questions but now needs empirical evidence. Empirical studies are being undertaken to establish *ELE* meets the criteria for validity and reliability and that the study demonstrates that the conceptual design is evidence-based.

References

- Alibali, M. W., Spencer, R. C., Knox, L., & Kita, S. (2011). Spontaneous gestures influence strategy choices in problem solving. *Psychological Science* 22, 1138-1144.
- American Psychological Association (2012). Ethical Principles of Psychologists. <http://www.apa.org/ethics/code/index.aspx#>. Accessed on February 03 2012
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 508–600). Washington: American Council on Education.
- Bellingham, L. (2008). Quality assurance and the use of subject level reference points in the UK. *Quality in Higher Education*, 14(3), 265-278.
- Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* [3rd edition]. Maidenhead, Berkshire: Society for Research into Higher Education & Open University Press.
- Biggs, J. B. & Collis, F. (1982). *Evaluating the quality of learning. The SOLO Taxonomy*. New York: Academic Press.
- Blessing, L.T.M. & Chakrabarti, A. (2009). *DRM, a Design Research Methodology*. Dordrecht: Springer.
- Bologna Working Group (2005). *A framework for qualifications of the European Higher Education Area*. Copenhagen: Ministry of Science, Technology and Innovation.
- Bransford, J., Brown, A., & Cocking, R.(Eds.). (2000). *How people learn*. Washington DC: National Academy Press.
- Chan, C. C., Tsui, M., Chan, M., & Hong, J. (2002). Applying the structure of the observed learning outcomes (SOLO) taxonomy on students' learning outcomes. *Assessment & Evaluation in Higher Education*, 27(6), 511-527.
- Chen, C. (2010). Teaching problem-solving and database skills that transfer. *Journal of Business Research*, 63, 175-181.
- Cheniaux, E, Landeira-Fernandez, J., Versiani, M. (2009) The Diagnoses of Schizophrenia, Schizoaffective Disorder, Bipolar Disorder and Unipolar Depression: Interrater Reliability and Congruence between DSM-IV and ICD-10. *Psychopathology* 2009;42:293-298

- Cicchetti, D. V. (1976). Assessing interrater reliabilities for rating scales. *British Journal of Psychiatry*, 129, 452-456.
- Cosgrove, R. (2011). Critical thinking in the Oxford tutorial. *Higher Education Research & Development*, 30(3), 343-356.
- COTAN [Committee on Test Affairs Netherlands] (2011) System of assessing the quality of tests of Psychologists/NIP] <http://www.psynip.nl>. Accessed 10 October, 2011.
- Creswell, J. W. (2007). *Qualitative inquiry & research design*. London: Sage.
- Cronbach, L. J. (1992). Four psychological bulletin articles in perspective. *Psychological Bulletin*, 112(3), 389-392.
- European Association of Quality Assurance (ENQA) (2009). *Standards and Guidelines for Quality Assurance in the European Higher Education Area*. 3rd edition. Available online at: http://www.enqa.eu/pubs_lasso. Accessed October 29, 2009.
- Ewell, P. (2010). Twenty years of quality assurance in higher education. *Quality in Higher Education*, 16(2), 173-176.
- Field, A. (2009). *Discovering statistics using SPSS* [3rd edition]. London: Sage.
- Harkness, J., Van de Vijver, F., & Mohler, P. (2003). Comparative research. In: *Cross-cultural survey methods*. New Jersey: John Wiley.
- Harris-Huermann, S. (2008). Evaluators of higher education in Germany. *Quality in Higher Education*, 14(1), 55-66.
- Hattie A.C. (2009) Visible learning. A synthesis of over 800 meta-analysis relating to achievement. New York: Routledge.
- Kane, M.Y. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* [4th edition] (pp. 17-64). Westport, CT: Praeger.
- Kek, M. & Huijser, H. (2011). The power of problem-based learning in developing critical thinking skills. *Higher Education Research & Development*, 30(3), 329-341.
- Konrad, H., Leynse, F., Crochet, M., Surssock, A., Campbell, C., Dill, D., Neetens, S. & Hover, C. (2007). *Report of the Committee for the Review of the Accreditation Organisation of the Netherlands and Flanders (NVAO)*. European Association of

- Quality Assurance (ENQA). http://www.enqa.eu/pubs_review.lasso. Accessed 29 October, 2009.
- Krathwohl, D., Bloom, B. & Masia, B. (1974). *Taxonomy of educational objectives*. New York: David McKay.
- Nusche, D. (2008). *Assessment of learning outcomes in higher education*. Organisation for Economic Co-operation and Development (OECD), Directorate for Education. Available online at: <http://www.oecd.org/dataoecd/13/25/40256023.pdf>. Accessed 5 July, 2010.
- Rexwinkel, G.B., Haenen, J.P.P. & Pilot, A. (2012b). Evaluating the level of degree programmes in higher education: the case of Nursing. *Assessment & Evaluation in Higher Education*. DOI: 10.1080/02602938.2012.751962.
- Robinson, S. (2011). Teaching logic and teaching critical thinking: revisiting McPeck. *Higher Education Research & Development*, 30(3), 275-287.
- Rutowski, L., & Rutowski, D. (2010). Getting it 'better'. *Journal of Curriculum Studies*, 42, 411-430.
- Sandahl, R., Rohlin, M., Waerness, M., Brennan, J., & Sjunesson, K. (2006). Evaluation of the Swedish National Agency for Higher Education European Association of Quality Assurance (ENQA) http://www.enqa.eu/pubs_review.lasso (2006). Accessed 29 October, 2009.
- Smith, T. (2003). Developing comparable questions in cross-national surveys. In J. Harkness, F. Van de Vijver & P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 69-92). New Jersey: John Wiley.
- Snijkers, G. (2002). Cognitive laboratory experiences. Doctoral Dissertation, Utrecht University. [Central Office of Statistics, Netherlands].
- Sternberg, R. (2003). What is an "expert student?" *Educational Researcher*, 32(8), 5-9.
- Westerheijden, D. Beerkens, E., Cremonini, L., Huisman, J., Kehm, B., Kovač, A., et al. (2009). The Bologna Process Independent Review. Available online at: http://ec.europa.eu/education/highereducation/doc/bologna_process/independent_review_1_detailed_rept.pdf. Accessed 9 February, 2011.
- Zimmerman, B. J. (2006). Development and adaptation of expertise: the role of self-regulatory processes and beliefs. In K. A. Ericsson, N. Charness, P. J. Feltovitch &

R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 705-742). New York: Cambridge University Press.

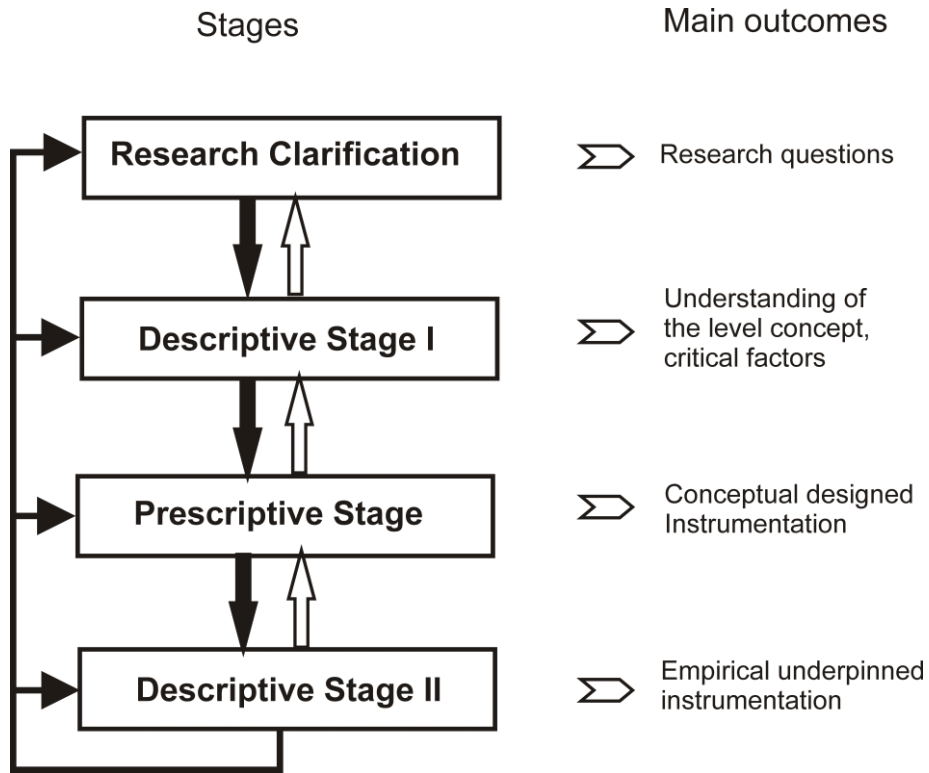


Figure 1 DRM Framework (Adapted from Blessing & Chakrabarti 2009)

<i>Vision</i>	Degree programmes in higher education should students learn to think on complex disciplinary problems at bachelor and master level. Higher education is the main place for students to develop disciplinary thinking and to attain a professional attitude.
<i>Analysis</i>	Consistent with the outlined vision the areas of the existing disciplines in the current curriculum, external surroundings, professional field, and academic field are analysed with the question ‘What findings and developments are important or urgent for the degree programme level?’
<i>Themes</i>	Building on the analysed findings and developments, themes are created that are relevant for the degree programme level. Substantiated by peer-reviewed literature, the themes are created with the knowledge, language and culture of the degree programme.
<i>Validation</i>	The resulting themes are validated by various stakeholders in the degree programme (face validity).
<i>Student aims</i>	The validated themes are specified with aims drawn up from the students’ perspective, that covers the themes, and express the appropriate degree of complexity in disciplinary thinking and professional attitude.
<i>Validation</i>	The student aims are tested on the extent the aims cover the themes (K)
<i>Items</i>	The student aims are transformed into items of a measuring instrument. The items express disciplinary thinking and professional attitude and are pretested for ambiguities, difficult syntax, mismatching questions and answers, overlapping categories, missing categories.
<i>Participants</i>	The participants from the measuring instrument’s target group should affirm the themes through the items in the real study and work setting.

Figure 2. Generic model of constituents addressing the validity

Table 1

SOLO taxonomy (Biggs & Collis, 1982) with examples of verbs (Biggs & Tang, 2007)

<i>SOLO-level and description</i>	<i>Verbs</i>
5. <i>Extended abstract</i> . There is recognition that the given example is an instance of a more general case. Hypotheses about examples <i>not</i> given are entertained and the conclusions are held open.	Reflect, create, compose, invent, make an original case, theorise, hypothesise, generalise, generate
4. <i>Relational</i> . Most or all of the evidence is accepted, and attempts are made to reconcile. Conflicting data are placed into a system that accounts for the given context. <i>Transitional</i> . There is a hint that closure, or a firm conclusion is not inevitable. There is a suggestion that a relating principle might account for the situation, but this is not spelled out.	Apply, integrate, analyse, explain, conclude, review, argue, transfer, make a plan, debate, construct, solve a problem
3. <i>Multi-structural</i> . Several consistent aspects of the data are selected, but any inconsistencies or conflicts are ignored or discounted so that a firm conclusion is reached. <i>Transitional</i> . Any inconsistencies are noted: Several aspects are recognisable but the student is unable to reconcile them.	Classify, describe, report, discuss, illustrate, select, compute, sequence, outline, separate
2. <i>Uni-structural</i> . An answer is based on only one relevant aspect of the presented evidence so that the conclusion is limited and probably dogmatic. <i>Transitional</i> . An attempt to handle two aspects of the evidence is made, but they may be inconsistent and hence no firm conclusion is reached.	Write, label, identify, recognise, count, define, find, match, memorise, quote
1. <i>Prestructural</i> . Student avoids the question (denial), repeats the question (tautology), a firm closure is based on transduction. <i>Transitional</i> . Student attempts to answer the question but only partially grasps a significant point.	Using tautology to cover lack of understanding, show little evidence of relevant learning.

Table 2

Attitude development scheme

Attitude Level and description

-
5. *Characterising*. At the highest level the student is able to place most aspects of the professional attitude consistently into a hierarchical control of his own behaviour. He is able to behave consistently with the aspects of the professional attitude.
 4. *Organising*. The student processes more than one aspect, and can organise these into a system without inconsistencies within the given system. He is able to relate aspects of the professional attitude and integrate these in a system. The self-regulation cycle(s) become more effective.
 3. *Valuing*. The student demonstrates that the required aspects of the professional attitude are becoming observable, and develops to consistency. The student understands the underlying meaning of the attitudinal aspects, although the student's responses are still inconsistent with the attitude he has identified,
 2. *Responding*. He accepts and tends toward the intended behaviour. Above that, he concentrates on the cycle(s) of self-regulation. He orients on goal, task, self-efficacy and outcome expectations to improve his behaviour through self-control and self-instruction. In the self-reflection phase the student determines whether he is mastering an aspect.
 1. *Receiving*. The student becomes aware of the aspects of the professional attitude. He does not yet feel the need for consistency of all the aspects. He also becomes aware of self-regulation cycle(s).
-