Evaluation Instrument for Engineering Modules and Courses

Elena Mäkiö-Marusik, Freeha Azmat, Bilal Ahmad, Robert Harrison
Warwick Manufacturing Group (WMG)
University of Warwick
Coventry, UK
(E.Makio-Marusik, f.azmat.1, B.Ahmad, Robert.Harrison)@warwick.ac.uk

Armando Walter Colombo
Faculty of Technology
University of Applied Sciences Emden/Leer
Emden/Leer, Germany
Armando.Colombo@hs-emden-leer.de

***Abstract*— Engineers of today are required to fulfil the growing demand of interdisciplinary skills required by society and industry. They are expected to possess not only profound disciplinary knowledge and skills, but also a range of methodical, social and personal competencies. A number of teaching modules have been delivered that aim to enhance those competencies in engineering students To evaluate quality of engineering modules an instrument is developed. This instrument measures acquired competencies, quality of the teaching process and settings. This paper presents the evaluation instrument and reports on its validity and reliability.**

Keywords—Engineering education; Evaluation of engineering modules and courses; Evaluation instrument; Assessment of competencies

I. Introduction

Society and businesses are placing constantly growing demands on today's engineers. To develop complex systems, e.g. of Industry 4.0 and Internet of Things, engineers are expected to possess profound disciplinary and interdisciplinary knowledge and skills combined with a broad range of methodical, social and personal competencies [1]. Those competencies, further called key competencies (also named as generic competencies, generic skills, or transferable skills in the literature), are not sufficiently addressed during engineering education and therefore are not developed well enough in engineering graduates [2] [3]. Academic institutions are responsible to society, industry, and engineering students to develop competencies required for engineering work and career. Perceiving their responsibility, academic institutions have developed a number of modules and courses that use novel didactical approaches and are aimed to enhance key competencies in engineering students, for instance [4] [5] [6] [7].

To address the development of a broad variety of competencies required in engineering graduates, the task-centric holistic agile teaching approach T-CHAT has been developed [8]. This approach has been used by delivering the teaching module "Sustainable Energy Systems" at Warwick Manufacturing Group (WMG), the University of Warwick. To validate a module delivered using this teaching approach, an appropriate evaluation methodology is required.

A number of module evaluation methodologies are reported in literature. For example, the modules reported in [4] [5] [6] [7] have been evaluated to assess developed knowledge, skills and competencies of students. However, different authors use different evaluation instruments which are tailored to the specific needs of these modules. Moreover, validity and reliability of those instruments have often not been tested.

Some evaluation instruments have been found in the literature [9] [10] [11] [12]. Though these evaluation instruments incorporate parameters for assessing competencies, they do not address the specific competence requirements of engineering modules. Furthermore, items which measure the general quality of a module or course are missing.

Therefore, a need has been perceived to develop an instrument for evaluation of engineering modules/courses that 1) combines evaluation of the module/course with assessment of developed competencies, 2) addresses the specifics of engineering education, and 3) can be used across a range of engineering module/course.

This paper is structured as follows: section II presents the background knowledge about competencies developed by modules in engineering education and how they may be evaluated. Section III describes how an evaluation questionnaire has been developed and which items it has. Section IV presents a short analysis of the validity and reliability of this evaluation instrument. Section V rounds up the paper with conclusions and highlights few outlooks.

II. Background

As aforementioned, an important learning outcome of engineering education is the development of competence in engineering graduates. This is supported by the competence-oriented and coordinated design of learning activities and assessment tasks [13]. Constructive alignment of these three elements - learning outcomes, activities and assessment tasks - forms a basis for development of modules. According to [9], quality of modules is assessed on the following criteria, or dimensions:

- learning outcomes of the module,
- learning/teaching process, and
- learning/teaching settings.

Learning outcomes determine the knowledge, skills and competencies that students may acquire in the course of a module. Learning/teaching process consists of a sequence of learning activities and assessment tasks carried out in the course of a learning module to achieve the learning outcomes. The dimension "learning/teaching process" means organization of learning/teaching, for instance structure of module, commitment of teacher. The dimension "settings" concerns learning/teaching facilities, for instance rooms and the number of participants. These three dimensions have been included in the evaluation instrument under development.

The learning/teaching process and settings are usually evaluated in form of a survey at the end of the module [9]. Assessing competencies, and especially key competencies, is acknowledged as difficult to measure in the literature [9] [14] and depends on several factors, for example on instructors (may have different degree of experience) and on student populations (may have different experience and motivation to learn) [15]. It cannot be measured through formal tests. Rather than measuring competencies directly via performance tests, students' self-assessment survey is the most effective way to assess acquired competencies [9] [16]. Such student survey measures competence that students developed during a module or a student team project. Self-assessment is often criticized for being subjective and consequently not as reliable as objective measures [17]. However, some research underpins the validity of self-assessment surveys. For instance, studies [18] [19] revealed that there is an association between self-assessments and the results of achievement tests. Self-reported responses of groups of students are rather valid and reliable than those of individual students [20].

Both disciplinary and key competencies will be assessed in the evaluation instrument under development. While disciplinary competencies are subject-specific, key competencies are general and can be classified according to [21] in four categories:

- social competence (e.g. the ability to communicate and collaborate),
- personal competence (e.g. responsibility, self-esteem, leadership),
- systematic competence (e.g. problem-solving and analytical skills), and
- general competence (e.g. project management, information technology).

This raises a question that which competencies need to be assessed by the evaluation instrument under development? Based on the literature review on required competencies and existing gaps in engineering graduates (see [22] [2]), as well as on the criteria set by Accreditation Board for Engineering and Technology (ABET) [23] and European Network for Accreditation of Engineering Education [24], the required competencies are summarised in Table I. The instrument under development should make it possible to assess these competencies.

TABLE I. REQUIREMENTS ON COMPETENCIES OF ENGINEERING GRADUATES

Competence	Reference
Disciplinary competence	
High level cognitive skills	[22] [2] [23] [24]
Engineering practice	[22] [2] [23] [24]
Social competence	
Collaboration	[22] [2] [23] [24]
Communication	[22] [2] [23] [24]
Presentation	[22]
Writing	[22] [2]
Personal competence	
Lifelong learning	[22] [23] [24]
Creative thinking	[22] [2]
Leadership	[2]
Innovation	[2]
Systematic competence	
Critical thinking	[22]
Analytical skills	[22] [2]
Problem-solving skills	[22] [2] [23]
Cross-disciplinary thinking	[22]
Innovation	[22]
Entrepreneurship, successful transferring plans into reality	[22] [2]
General competence	
Project management	[22]

III. DEVELOPMENT OF EVALUATION QUESTIONNAIRE

According to [25], three guidelines should be followed for the design of student evaluation instruments:

- the list of items must be short to exclude cognitive biases and "carelessness due to student boredom";
- the items must be agreed by experts; and
- the items must be "susceptible to student observation and judgement".

To develop the required evaluation instrument, a literature review on the existing questionnaires has been conducted. An overview of nine published student questionnaires used for evaluation of teaching in German universities is given in [9]. Other four instruments to measure teaching effectiveness are presented in [25]. The common feature of all these instruments is that they contain only questions on teaching process and settings, but not on outcomes acquired. However, there are some instruments in the literature that measure acquisition of competence. For instance, in [9] a questionnaire for assessment of competencies, for example, disciplinary competence, methods expertise, presentation, communication, collaboration and personal competencies are included. This instrument however does not fulfil requirements on competencies of engineering modules. A questionnaire for measuring interdisciplinary competence used in [12] is too specific. An evaluation questionnaire which is used in [26] and measures four scales – knowledge, skills, social competencies and course organisation, is extensive, but misses learning/teaching scale. To develop a practical and appropriate evaluation instrument, the knowledge and insights gained from these studies have been used.

Furthermore, some recommendations for designing of self-assessment questionnaires have been found in the literature.

For instance, some rules for conducting a self-assessment survey and formulating questions are summarized in [12]. First of all, the respondents need to be aware of the information requested in the survey. Questions should be phrased “clearly and unambiguously” and refer to recent activities. The respondents should think that “the questions merit a serious and thoughtful response” and that their answer does not violate their privacy. According to [9], survey questions must only have first-person statements. Self-assessment questions for key competencies must measure acquisition of those competencies in the course of a teaching module, and not their current level. Hence such words as “better” and “more” should be used formulating questions. Questions about disciplinary competencies, in contrast, can ask about the current level of those.

Marsh claims in [25] that an evaluation instrument is typically constructed by a logical analysis of the content of what should be evaluated and by a review of evaluation literature in this field. Empirical factor analysis to test the proposed scales supplements this theoretical process. To construct the evaluation instrument, the field of competences and methods of their evaluation has been analysed, literature with the existing questionnaires has been reviewed, and the aforementioned recommendations have been taken into account. In the questionnaire three quality dimensions of a teaching module– outcome, process, and settings – have been included. Based on these quality dimensions, the aforementioned classification of key competencies, and the list of the required competencies in Table I, a factor structure of the questionnaire with the following scales (in other words factors or latent variables) is developed:

1. disciplinary competence,
2. social competence,
3. systematic competence,
4. learning process, and
5. settings.

Averaged student ratings may be interpreted according this factor structure.

The revised Blooms’ taxonomy [27] has been used to formulate items for self-assessment of disciplinary competencies. Some items on social and systematic competencies have been derived from the questionnaires introduced in [9] and [26]. Items about learning/teaching process and settings of a teaching module have been taken from the module feedback questionnaire used at WMG, the University of Warwick. For the personal competencies leadership and innovation no items are included because it is difficult to qualitatively assess these competencies. Table II lists the coded items along with the assessed competence and source. All items in the evaluation questionnaire will be measured on a 5−point Likert−type scale: Definitely agree, Mostly agree, Neither agree nor disagree, Mostly disagree, Definitely disagree.

TABLE II. LIST OF ITEMS FORTHE COMPETENCE EVALUATION INSTRUMENT

Code	**Item**	**Competence to be assessed**	**Source**
DC	**Disciplinary competence**		
Q1	Due to this module, I understand the basic definitions of the subject.	Bloom’s level 2 (Understand)	[28]
Q2	Due to this module, I understand the fundamental problems of the subject.	Bloom’s level 2 (Understand)	[28]
Q3	Due to this module, I am able to choose the adequate methods to the problems of this field.	Bloom’s level 3 (Apply)	[28]
Q4	I am able to use basic theoretical knowledge and practical skills in the subject.	Bloom’s level 3 (Apply)	[28]
Q5	I am able to analyze solutions and processes of the subject.	Bloom’s level 4 (Analyze)	
Q6	I am able to argue and evaluate the given problems and solutions of the topic.	Bloom’s level 5 (Evaluate)	[28]
Q7	I am able to compare and find significant connections and correlations in the field.	Bloom’s level 5 (Evaluate)	[28]
Q8	I am able to formulate solutions using the methods, techniques and tools of the subject.	Bloom’s level 6 (Create)	[28]
SoC	**Social competence**		
Q9	Due to this module it is easier for me to express my own opinions.	Communication	[9]
Q10	Due to this module I make my verbal contributions in more comprehensible language.	Communication	[9]
Q11	Due to this module it is easier for me to ask when I have not understood something.	Communication	[9]
Q12	I participated in the work planning within the team during this module.	Collaboration	[9]
Q13	I contributed to the assignment of tasks within the team during this module.	Collaboration	[9]
Q14	Due to this module I can better hold a presentation.	Presentation	[9]
Q15	Due to this module I can better write technical texts.	Technical writing	[9]
SyC	**Systematic competence**		
Q16	Due to this module I can better critically question and evaluate new ideas/things.	Critical thinking	
Q17	Due to this module I can better think across technical and non-technical considerations, can better see things from different perspectives.	Cross-disciplinary thinking	
Q18	Due to this module I can work more systematically and logically, can better collect, visualize and analyze information.	Analytical skills	
Q19	Due to this module I can better identify and develop new things at my workplace/in my own projects.	Successful transferring plans into reality, creativity	
Q20	Due to this module I can better solve problems of different nature that I encounter at my workplace/in my own projects.	Problem-solving skills	
Q21	Due to this module I can better manage my future projects as well as projects at my workplace.	Project management	

Q22	Due to this module I can better find and apply information about methods, techniques and tools needed to solve an issue.	Lifelong learning, Self-directed learning	
LP	**Learning process**		
Q23	The objectives are clear.		
Q24	The content is appropriate.		
Q25	The content is interesting.		
Q26	The information in this module is appropriate for me / my company/workplace.		
Q27	Staff are good at explaining things.		
Q28	Staff make the subject interesting.		
Q29	Staff are enthusiastic about what they taught.		
Q30	The module is intellectually stimulating.		
Q31	I am happy with the pace of learning.		
Set	**Settings**		
Q32	The timetable worked efficiently.		
Q33	The module was well organized and ran smoothly.		
Q34	I have been able to contact staff when I needed to.		
Q35	I have received sufficient advice and support for the module.		
Q36	The resources for this module are sufficient for my needs.		
Q37	Notes support the learning.		
Q38	I have been able to access resources when I needed to.		

IV. VALIDITY AND RELIABILITY OF EVALUATION ISTRUMENT

Before using the proposed instrument for evaluation of modules it is required to test whether it is valid and reliable. According to [29], an instrument is valid when it "measures what it purports to measure". An instrument is reliable when its measurements are consistent. A reliable instrument is not necessarily valid for a particular purpose. For the proposed instrument the following tests have been conducted:

1. content and construct validity tests, and
2. internal consistency and intra-class correlation tests (for reliability).

To test construct validity, internal consistency and intra-class correlation, methods of social statistics have been applied. A sample of students has been taken who completed technical modules at the University of Applied Sciences Emden/Leer, Germany in the winter term 2018/2019. 135 questionnaires have been filled out, 36 of them with missing values. The sample size is between poor (100) and fair (200) [30] that brings some limitation in analysis.

A. Descriptive Statistics

Descriptive statistics of the items indicate a response pattern of the individual questions (see Table III). Item mean, standard deviation, skewness, kurtosis are reported. Most of the data is skewed left (the left tail is longer) with skewness coefficients (sk) between -1 and -0.5 that indicates a moderately skewed distribution. Kurtosis (k) measures outliers in a data set with a larger kurtosis indicating a serious outlier problem. In the current data set kurtosis for all items except of Q2 is between -2 and +2. The values of skewness and kurtosis are acceptable in order to prove normal univariate distribution [31].

TABLE III. ITEM CHARACTERISTICS OF THE EVALUATION INSTRUMENT

Code	m	std	sk	k	r_{it}	d_{it}	fl
DC	**Disciplinary competence**						
Q1	4.00	.810	-.939	1.347	.547	.963	.55
Q2	3.96	.799	-1.102	2.295	.525	.963	.43
Q3	3.67	.859	-.685	.534	.593	.963	.61
Q4	3.99	.815	-.658	.180	.489	.963	.66
Q5	4.01	.768	-.414	-.200	.456	.964	.66
Q6	3.92	.804	-.663	.764	.367	.964	.75
Q7	3.73	.839	-.471	.173	.533	.963	.80
Q8	3.87	.725	-.532	.483	.522	.963	.79
SoC	**Social competence**						
Q9	3.33	1.182	-.434	-.386	.680	.962	.66
Q10	3.42	1.113	-.523	-.114	.777	.962	.69
Q11	3.48	1.293	-.541	-.658	.686	.962	.62
Q12	3.06	1.446	-.141	-1.206	.712	.962	.79
Q13	3.01	1.456	-.140	-1.231	.664	.963	.84
Q14	2.60	1.471	.278	-1.296	.685	.963	.90
Q15	2.68	1.339	.080	-1.223	.744	.962	.89
SyC	**Systematic competence**						
Q16	3.15	1.173	-.468	-.600	.697	.962	.70
Q17	3.28	1.107	-.534	-.255	.742	.962	.74
Q18	3.40	1.184	-.489	-.534	.666	.963	.70
Q19	3.23	1.222	-.366	-.705	.752	.962	.86
Q20	3.24	1.142	-.446	-.364	.704	.962	.76
Q21	3.11	1.204	-.316	-.621	.763	.962	.81
Q22	3.47	1.125	-.662	-.016	.727	.962	.81
LP	**Learning process**						
Q23	4.01	.958	-.969	.602	.669	.963	.74
Q24	3.93	.959	-.932	.730	.560	.963	.65
Q25	3.90	1.014	-.895	.476	.626	.963	.66
Q26	3.65	1.105	-.691	-.005	.659	.963	.49
Q27	3.39	1.288	-.514	-.826	.684	.962	.69
Q28	3.38	1.227	-.488	-.681	.709	.962	.69
Q29	3.85	1.066	-.984	.547	.598	.963	.66
Q30	3.75	1.087	-.826	.249	.708	.962	.73
Q31	3.61	1.233	-.737	-.370	.721	.962	.84
Set	**Settings.**						
Q32	3.62	1.144	-.683	-.264	.530	.963	.64
Q33	3.55	1.128	-.380	-.842	.665	.963	.71
Q34	3.95	.958	-.687	.089	.390	.964	.39
Q35	3.80	.922	-.438	-.571	.528	.963	.62
Q36	3.79	1.055	-.916	.500	.634	.963	.69
Q37	3.67	1.196	-.797	-.221	.687	.962	.62
Q38	3.94	.925	-.778	.466	.408	.964	.41

m = mean, std = standard deviation, sk = skewness, k = kurtosis, r_{it} = Corrected Item – Total Correlation, d_{it} = Cronbach's α if item deleted, fl = Factor Loadings

B. Validity Tests

Content and construct validity tests have been conducted to test validity of the proposed evaluation instrument. Content validity ensures that an evaluation instrument includes only required items of a particular construct domain. The approach to establish content validity begins with an extensive literature review followed by an expert evaluation. To develop the proposed instrument, a literature review has been conducted on the existing instruments that evaluate competencies. The items selected in this review have been validated by experts of University of Warwick. Thus, the research and development process for survey items contributes to the construct validity of the measure.

To test the construct validity a confirmatory factor analysis (CFA) has been conducted using statistical software Stata. The aim of this analysis is to demonstrate that scores on scales (in other words factors or latent variables, i.e. disciplinary competence, social competence, systematic competence, learning process, and settings) covary with the scores on the corresponding observed variables. Thus, factor analysis tests whether students are able to differentiate among different scales of the evaluation instrument and whether the empirical scales confirm the aspects that the instrument is developed to measure.

A robust Maximum likelihood estimator has been used for CFA. This estimator delivers Chi^2 test which is a conservative measure of the model accuracy and is very sensitive to sample size. Therefore the model fit is evaluated based on the relation of χ^2 statistic and degrees of freedom, as well as other global fit indices will be calculated: the Comparative Fit Index (CFI), the Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Factor loadings of items are examined. Maximum likelihood estimation has been first conducted on the observations that have only complete data. RMSEA of the model however did not satisfy conventional cut-off values. The reason of that may be the small sample size (135 observations, 36 of them with missing values). Maximum likelihood estimation with missing values has been conducted that resulted in model confirmation. However, SRMR could not be reported because of missing values.

Correlations between the scales of the instrument (see Table IV) are low, except of moderate correlation (0.62) between scales systematic and social competence. In contrast, the vast majority of factor loadings (fl) of the individual items is between 0.60 and 0.90 (see Table III). Only single observed variables (13%) are between 0.39 and 0.55. To sum up, all latent constructs are well constituted by their observed variables and explain their variances well.

TABLE IV. INTERCORRELATIONS OF SCALES

Scale	DC	SC	SyC	LP	Set
DC Disciplinary competence	1.00				
SC Social competence	.16	1.00			
SyC Systematic competence	.23	.62	1.00		.
LP Learning process	.25	.31	.43	1.00	
Set Settings	.22	.34	.42	.47	1.00

Item-total statistics of the individual items within each scale are presented in Table III. Corrected Item – Total Correlation coefficients are all $r_{it} > 0.3$. This indicates that each item correlates with all other items, so the scale performs homogeneous measurement.

The overall test of fit (the χ^2 statistic) resulted in $\chi^2 = 905.980$ by 624 degrees of freedom. The ratio of χ^2 to the degrees of freedom is 1.45 and satisfies the cut-off criterion to be less than 2.00 [32] what indicates model fit. While comparative fit index (CFI) = 0.918 is under the recommended limit of 0.95, RMSEA = 0.058 is between the recommended range 0.05 and 0.08 that indicates fair fit [33].

C. *Reliability Tests*

Reliability of an instrument measures the relative consistency of responses among raters [34], so indicating the relative lack of random error in student answers. Reliability determines consistency or stability of an instrument. One type of reliability called the intra-class correlation (here represented by ICC(2)) "provides an estimate of the reliability of the group means" This ICC(2) indicates difference in between-group and within-group variances and is estimated from the ICC(1) ("an index of interrater reliability, the extent to which raters are substitutable" [34]) using the Spearman-Brown formula with regard to the group size [34]. ICC(2) is high by small within-group variance relative to between-group variance. Intra-class correlation of the instrument ICC(2) = 0.912 for the entire sample what means very high interrater-reliability. ICC(2) = 0.530 when based on 10 students, ICC(2) = 0.656 when based on 20 students, and ICC(2) = 0.708 when based on 30 students. Consequently, any summary report based on fewer than 30 responses needs to be used with caution.

Another type of reliability, the internal consistency, provides information about the extent to which the items of a factor measure the same construct. The internal consistency of the entire instrument is very good, with Cronbach's $\alpha = 0.964$. The internal consistency of the five scales of the instrument has been calculated (see Table V): the Cronbach's alphas vary between 0.832 and 0.930 what indicates a good to very good internal consistency.

Item-total statistics of the individual items are presented in Table III. The coefficients "Cronbach's α if item deleted" (d_{it}) are equal or slightly below of Cronbach's $\alpha = 0.964$ of the instrument. Corrected Item – Total Correlation coefficients are all $r_{it} > 0.3$. This indicates that each item correlates with all other items, so the scale performs homogeneous measurement.

TABLE V. RELIABILITY OF THE SCALES

Scale	Number of Items	Cronbachs α
DC Disciplinary competence	8	.875
SC Social competence	7	.930
SyC Systematic competence	7	.919
LP Learning process	9	.895
Set Settings	7	.832

D. *Limitations of the Study*

This study has several limitations that should be considered when interpreting the results. First, the sample used to test content validity and reliability is between poor (100) an fair (200) [30]. Due to the limited resources content validity has not been thoroughly conducted, as well as criterion validity check have not been performed. Intra-class correlation test showed that any summary report on data which was collected using the proposed instrument and have fewer than 30 responses needs to be used with caution.

V. Conclusions and Outlook

To address the development of a broad variety of competencies required in engineering graduates, the task-centric holistic agile teaching approach T-CHAT has been developed by the authors. T-CHAT has been used by delivering the teaching module "Sustainable Energy Systems" at WMG, the University of Warwick. This paper introduces an evaluation instrument that has been developed by the authors in order to validate the addressed module. Evidence supports the construct validity of the five scales of the instrument. Reliability of the instrument is very good. This questionnaire has been first used for evaluation of the teaching module "Sustainable Energy Systems". The authors recommend that this instrument may be used for evaluation of other engineering modules.

An important direction for future improvement of this instrument, is to check content validity of the instrument through interviews with students. These interviews should especially be focussed on competencies they acquired and the comparison of the interview results with the outcomes of the qualitative survey. Such research would provide additional evidence of the validity of the instrument. The very successful results obtained till now make the authors enthusiastic in further applying both the T-CHAT teaching approach and the evaluation instrument in some modules of the graduate course "Industrial Informatics" at the University of Applied Sciences Emden/Leer, Germany.

References

[1] National Research Council, *Interim Report on 21st Century Cyber-Physical Systems Education*, The National Academies Press, 2015.

[2] S.A. Male, "Generic engineering competencies: A review and modelling approach," *Education Research and Perspectives*, vol. 37, 2010, p. 25.

[3] E. Mäkiö-Marusik, B. Ahmad, R. Harrison, J. Mäkiö, and A.W. Colombo, "Competences of Cyber Physical Systems Engineers – Survey Results," *Industrial Cyber-Physical Systems (ICPS), 2018 IEEE 1st International Conference on*, IEEE, 2018, p. xxx–yyy.

[4] G. Verbic, C. Keerthisinghe, and A.C. Chapman, "A Project-Based Cooperative Approach to Teaching Sustainable Energy Systems.," *IEEE Transactions on Education*, vol. 60, 2017, pp. 221–228.

[5] R.M. Lima, J. Dinis-Carvalho, R.M. Sousa, A.C. Alves, F. Moreira, S. Fernandes, and D. Mesquita, "Ten Years of Project-Based Learning (PBL) in Industrial Engineering and Management at the University of Minho," *PBL in Engineering Education*, Springer, 2017, pp. 33–51.

[6] N. Arana-Arexolaleiba and M.I. Zubizarreta, "PBL Experience in Engineering School of Mondragon University," *PBL in Engineering Education*, Springer, 2017, pp. 89–102.

[7] R.P. Raycheva, D.I. Angelova, and P.M. Vodenova, "Project-based learning in engineering design in Bulgaria: expectations, experiments and results," *European Journal of Engineering Education*, vol. 42, 2017, pp. 944–961.

[8] J. Mäkiö, E. Mäkiö-Marusik, and E. Yablochnikov, "Task-centric holistic agile approach on teaching cyber physical systems engineering," *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, IEEE, 2016, pp. 6608–6614.

[9] E. Braun, *Das Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen (BEvaKomp)*, Göttingen: V&R unipress, 2008.

[10] A. Frey and L. Balzer, "Der Beurteilungsbogen smk: Ein Messverfahren für die Diagnose von sozialen und methodischen Fähigkeitskonzepten," *Kompetenzdiagnostik–Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen*, 2005, pp. 31–56.

[11] H.W. Marsh, "SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of University teaching," *British journal of educational psychology*, vol. 52, 1982, pp. 77–95.

[12] L.R. Lattuca, D. Knight, and I. Bergom, "Developing a measure of interdisciplinary competence," *International journal of engineering education*, vol. 29, 2013, pp. 726–739.

[13] J.B. Biggs, *Teaching for quality learning at university: What the student does*, McGraw-Hill Education (UK), 2011.

[14] L.J. Shuman, M. Besterfield-Sacre, and J. McGourty, "The ABET 'professional skills'—Can they be taught? Can they be assessed?," *Journal of engineering education*, vol. 94, 2005, pp. 41–55.

[15] M.J. Prince and R.M. Felder, "Inductive teaching and learning methods: Definitions, comparisons, and research bases," *Journal of engineering education*, vol. 95, 2006, pp. 123–138.

[16] H. Schaeper, "Development of competencies and teaching–learning arrangements in higher education: findings from Germany," *Studies in Higher Education*, vol. 34, 2009, pp. 677–697.

[17] S.R. Porter, "Do college student surveys have any validity?," *The Review of Higher Education*, vol. 35, 2011, pp. 45–76.

[18] G.R. Pike, "The relationship between self reports of college experiences and achievement test scores," *Research in higher education*, vol. 36, 1995, pp. 1–21.

[19] G.R. Pike, "Limitations of using students' self-reports of academic development as proxies for traditional achievement measures," *Research in higher education*, vol. 37, 1996, pp. 89–114.

[20] J.F. Volkwein, L.R. Lattuca, P.T. Terenzini, L.C. Strauss, and J. Sukhbaatar, "Engineering change: A study of the impact of EC2000," *International Journal of Engineering Education*, vol. 20, 2004, pp. 318–328.

[21] H. Orth, *Schlüsselqualifikationen an deutschen Hochschulen: Konzepte, Standpunkte und Perspektiven*, Luchterhand, 1999.

[22] E. Mäkiö-Marusik, "Current trends in teaching cyber physical systems engineering: A literature review," *Industrial Informatics (INDIN), 2017 IEEE 15th International Conference on*, IEEE, 2017, pp. 518–525.

[23] Engineering Accreditation Commission, "Criteria for accrediting engineering programs," *Accreditation Board for Engineering and Technology Inc*, 2015.

[24] European Network for Accreditation of Engineering Education, "EUR-ACE Framework Standards for the Accreditation of Engineering Programmes," *Brussels*, 2008.

[25] H.W. Marsh, "Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research," *International journal of educational research*, vol. 11, 1987, pp. 253–388.

[26] E. Mäkiö-Marusik, J. Mäkiö, and J. Kowal, "Implementation of Task-Centric Holistic Agile Approach on Teaching Cyber Physical Systems Engineering," *AMCIS 2017 Proceedings: Proceedings of the Twenty-third Americas Conference on Information Systems, IS in Education, IS Curriculum, Education and Teaching Cases (SIGED)*, 2017, pp. 1–10.

[27] L.W. Anderson and D.R. Krathwohl, "A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives, abridged edition," 2001.

[28] E. Elena Mäkiö-Marusik, J. Mäkiö, and J. Kowal, "Validation of Task-centric Holistic Agile Approach on Teaching Cyber Physical Systems Engineering," *GOSPODARKA RYNEK EDUKACJA= ECONOMY MARKET EDUCATION*, vol. 18, 2017, pp. 5–17.

[29] N.J. Salkind, *Encyclopedia of measurement and statistics*, SAGE publications, 2006.

[30] B.G. Tabachnick and L.S. Fidell, *Using multivariate statistics*, Allyn & Bacon/Pearson Education, 2007.

[31] W.M. Trochim and J.P. Donnelly, "Research methods knowledge base," 2001.

[32] A. Beauducel and W.W. Wittmann, "Simulation study on fit indexes in CFA based on data with slightly distorted simple structure," *Structural Equation Modeling*, vol. 12, 2005, pp. 41–75.

[33] L. Hu and P.M. Bentler, "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives," *Structural equation modeling: a multidisciplinary journal*, vol. 6, 1999, pp. 1–55.

[34] P.D. Bliese, "Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis.," 2000.