Evaluation of competency methods in engineering education

A systematic review

Leandro Cruz, Mariana; Saunders, Gillian; Groen, Pim

DOI
10.1080/03043797.2019.1671810

Publication date
2019

Document Version
Final published version

Published in
European Journal of Engineering Education

European Journal of Engineering Education

ISSN: 0304-3797 (Print) 1469-5898 (Online) Journal homepage: https://www.tandfonline.com/loi/ceee20

Evaluation of competency methods in engineering education: a systematic review

Mariana Leandro Cruz, Gillian N. Saunders-Smits & Pim Groen

EUROPEAN JOURNAL OF ENGINEERING EDUCATION
https://doi.org/10.1080/03043797.2019.1671810

REVIEW

OPEN ACCESS

Evaluation of competency methods in engineering education: a systematic review

Mariana Leandro Cruz, Gillian N. Saunders-Smits and Pim Groen

Aerospace Structures and Materials Department, Delft University of Technology, Delft, Netherlands

ABSTRACT

The purpose of this systematic review is to evaluate the state-of-the-art of competency measurement methods with an aim to inform the creation of reliable and valid measures of student mastery of competencies in communication, lifelong learning, innovation/creativity and teamwork in engineering education. We identified 99 studies published in three databases over the last 17 years. For each study, purpose, corresponding methods, criteria used to establish competencies, and validity and reliability properties were evaluated. This analysis identified several measurement methods of which questionnaires and rubrics were the most used. Many measurement methods were found to lack competency definitions and evidence of validity and reliability. These show a clear need for establishing professional standards when measuring mastery of competencies. Therefore, in this paper, we propose guidelines for the design of reliable and valid measurement methods to be used by educators and researchers.

ARTICLE HISTORY
Received 28 August 2018
Accepted 15 September 2019

KEYWORDS
Transversal competencies; competency measurement; competency assessment; engineering education; systematic review

1. Introduction

Over the last 20 years, accreditation boards and educational stakeholders worldwide have emphasised the importance of integrating transversal competencies in engineering education curricula in order to prepare students for the engineering labour market (American Society for Engineering Education 1994; Engineering Accreditation Commission 2000; UNESCO 2010). Transversal competencies were first defined by Care (Care and Luo 2016) as 'skills, values and attitudes that are required for learners' holistic development and for learners to become capable of adapting to change' and are also known in the literature as employability skills (Markes 2006), generic skills (Bennett, Dunne, and Carré 2000), key competencies (Organisation for Economic Co-Operation Development 2005), non-technical skills (Knobbs and Grayson 2012), non-traditional skills (Crawley et al. 2007), professional skills (Shuman, Besterfield-Sacre, and McGourty 2005), soft skills (Whitmore and Fry 1974), transferable skills (Kemp and Seagraves 1995), and twenty-first century skills (Council 2013).

The growing emphasis on transversal competencies in engineering education has triggered the need to create robust methods that measure transversal competencies (Shuman, Besterfield-Sacre, and McGourty 2005). However, assessing students' level of mastery in transversal competencies is difficult, caused in part by a lack of consensus on the definition of the transversal competencies between the different engineering education communities, government bodies, and employers, and by what behaviours would exhibit mastery (Shuman, Besterfield-Sacre, and McGourty 2005). In addition, it is also difficult to assess transversal competencies independently, because they are

CONTACT Mariana Leandro Cruz M.LeandroCruz@tudelft.nl

often intertwined with the technical competencies (Shuman, Besterfield-Sacre, and McGourty 2005; Badcock, Pattison, and Harris 2010). These issues have hindered the development of competency measurement process.

This work is part of an Erasmus+ Knowledge Alliance known as the PREFER project that aims to improve the employability of future engineers. Within this project, we are developing curriculum elements that assist students in developing transversal competencies in communication, lifelong learning, innovation and teamwork. To evaluate the effectiveness of curriculum elements stimulating these competencies, we have reviewed the competency measurement methods present in engineering education literature.

With this review, we aim to inform the creation of reliable and valid measures of student mastery of competencies in communication, lifelong learning, innovation/creativity and teamwork in engineering education. To do so, we look at methods which are used 1) to evaluate course and programme effectiveness to enhance the quality of teaching and student learning, 2) to assess students' performance with the purposes of giving summative grading and/or formative feedback, and 3) to measure students' abilities in order to characterise student populations.

The following research questions were addressed in this review:

(1) What are the methods used to measure the competencies: communication, innovation/creativity, lifelong learning and teamwork?
(2) Are validity and reliability measured in the studies considered, and if so, which techniques are used?
(3) What is the purpose of the measurement used in the study?
(4) Which criteria are used to assess these competencies?

2. Background

This section provides the reader with the motivation for the selected competencies in study. This selection has been carried out using scientific and industry literature and within the confines and scope of the PREFER project.

The need to focus on transversal competencies in the engineering curricula was first highlighted in 1996 by McMasters and Matsch (McMasters and Matsch 1996) in the Boeing list of 'Desired Attributes of an Engineer'. This list required engineers to have *good communication skills: written, verbal, graphic, listening*, *ability to think both critically and creatively*, *curiosity and a desire to learn - for life*, and *profound understanding of the importance of teamwork* (McMasters and Matsch 1996).

Further emphasis on competencies such as communication, working in teams, and lifelong learning was given by the new ABET Engineering Criteria which came into effect in 2000 (Engineering Accreditation Commission 2000) and the Washington Accord (American Society for Engineering Education 1994). Similarly, in Europe after the Bologna process, which started in 1999, the European Network for Engineering Accreditation (ENAEE) has set these three competencies as an important part of engineering programmes.

A resulting engineering education initiative, called CDIO (Conceive, Design, Implement and Operate), which started in 1997 at MIT and is now a worldwide initiative, has developed a list of competencies which include *creative thinking*, *curiosity and lifelong learning*, *multidisciplinary teamwork*, and *communications* (Crawley et al. 2007).

In summary, we chose to limit ourselves to the competencies of communication, teamwork and lifelong learning, as the comparison of the competencies present in all the previously mentioned literature that showed agreement on the importance of these three competencies.

A fourth competency, innovation/creativity, was added within the framework of the PREFER project and was taken from the list of 'Great Eight Competencies' (Bartram 2005), a validated tool available and used in this project. This competency was found to be important based on the outcomes of a large industry consultation by another PREFER project partner (Craps et al. 2018). Considering the challenges of technology in the future, this competency is acknowledged essential for engineering students not only by the PREFER project but also by the wider engineering education community (Badran 2007; Crawley et al. 2007; Cropley 2015; Kamp 2016).

3. Methods

In this section, we describe the data collection methods used to carry out the systematic review (summarised in Figure 1) and report on the characteristics of the studies found.

3.1. Data collection

This review has been carried out based on the methods outlined in the practical guide on systematic review of Petticrew and Robert (Petticrew and Roberts 2006). Following this method, first, the research questions were framed, as stated in the section of the introduction. Next, the databases were chosen and the research terms defined. The research was carried out in October 2017 using three databases: ERIC (education indexes), Scopus (science, technology, medicine, social sciences, and art and humanities indexes), and Web of Science (sciences, arts, and humanities indexes). The following keywords: *communication*, *innovation*, *creativity*, *lifelong learning*, *life-long learning*, *teamwork*, or *collaboration*, in combination with *measure*, *assess*, *method* or *evaluate* and *engineering* were used in each of the three databases. In addition, controlled library terms (see the PRISMA diagram in Figure 1) were used after applying the keywords to filter the relevant studies. The research was limited to English language studies in peer-reviewed literature, scientific journals, and conference proceedings from 2000 to 2017. The choice of the year 2000 as the starting point reflects the introduction of the ABET criteria for engineering programmes in that year (Shuman, Besterfield-Sacre, and McGourty 2005). Within these parameters, 332, 391, and 349 studies were identified in Scopus, Web of Science, and ERIC, respectively. From these studies, eighty-five duplicates were removed, resulting in 987 studies to be considered.

The third step of the method was to formulate the inclusion and exclusion criteria. To be included, the study:

- Was performed on engineering students in higher, tertiary and postsecondary education. Studies on primary and secondary education, training of practising engineers, and non-engineering programmes were excluded.
- Looked at at least one of the selected competencies: communication, innovation/creativity, lifelong learning and teamwork.
- Reported on methods used to measure students' performances (i.e. grading and feedback), to evaluate course and programme outcomes, and to measure students' abilities in non-related courses.
- Reported its aims and research questions, contained an adequate description of the data (country, participants, etc.), and provided answers to the research questions.

The first author examined the titles and abstract content of the studies found against the first two criteria. Then, the same author scanned the full texts (110 studies) against the last two requirements. Studies that did not fulfil the criteria were removed from this study.

From this analysis, 99 suitable studies were identified and managed using an EndNote™ citation database.

To answer the research questions, data about the measurement criteria, the methods used to measure each competency and the purpose of the measurement (1- students' performance for formative and summative assessment, 2 - evaluation of course/programme effectiveness and 3 -

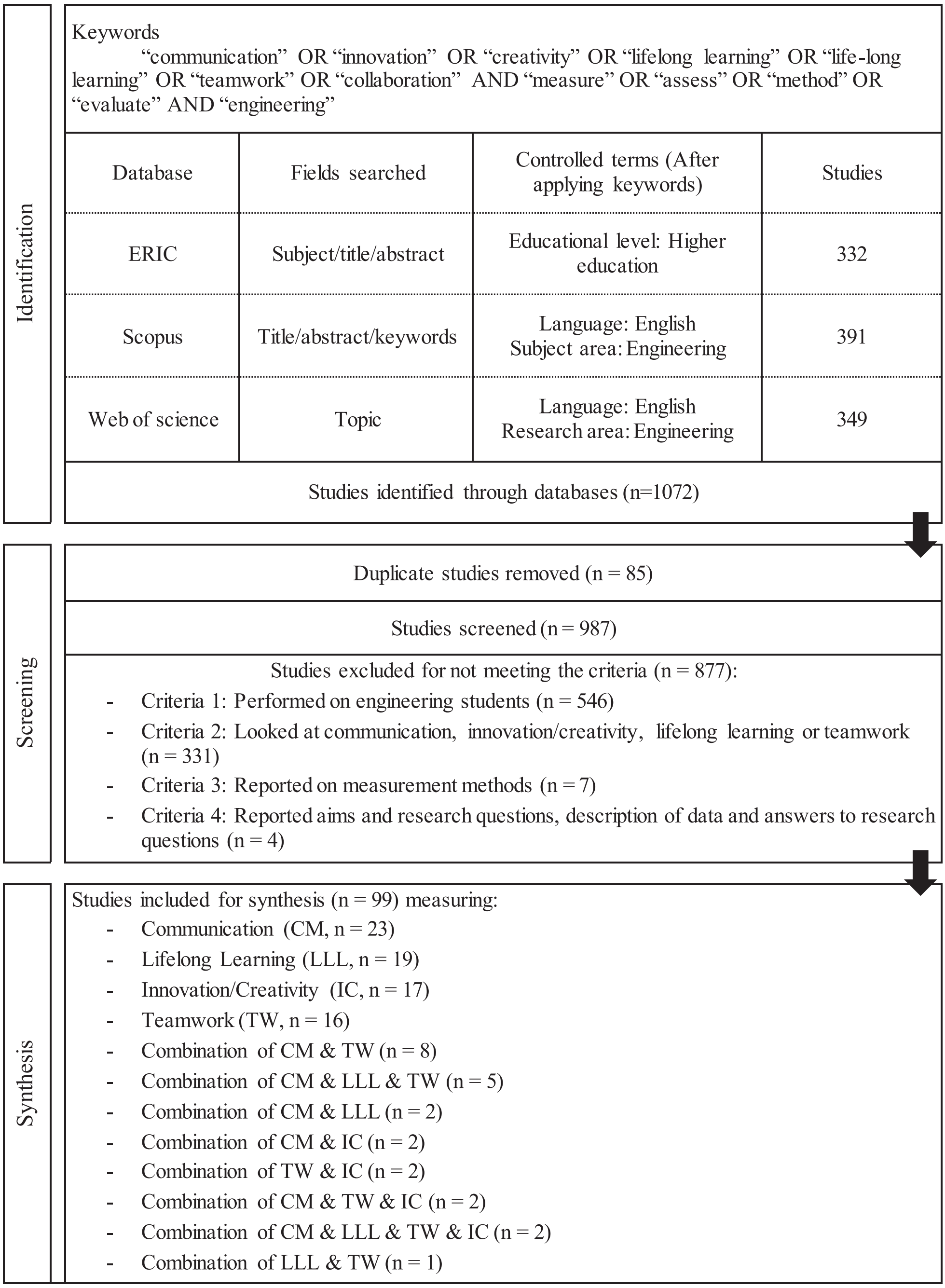

Figure 1. PRISMA flow diagram of the systematic review process.

characterisation of students' abilities) were extracted. In addition, the first author screened the studies to search for the use of the main types of validity and reliability measurements, as recommended by Cohen (Cohen, Manion, and Morrison 2007): content validity, construct validity, reliability as stability, reliability as equivalence and reliability as internal consistency. These data were recorded on a data sheet.

3.2. Study characteristics

When looking at the characteristics of the studies, only 17% of the studies were published between 2000–2009, compared to 83% published between 2010–2017 (Figure 2). The analysis of the geographical spread of the studies shows that the most studies (64%) on competency measurement originated in North America, followed by Europe (19%), South America (7%), Asia (5%), Australia (3%), and Africa (1%). Moreover, 75% of the studies looked at only one competency (see Figure 1). Only 2% of the studies (Moalosi, Molokwane, and Mothibedi 2012; Narayanan 2013) looked at all four competencies. Communication was the competency which was most frequently studied (44% of the studies), followed by teamwork (36%), lifelong learning (29%) and innovation/creativity (25%).

4. Results

The findings of the systematic review are structured to address the research questions. Firstly, the type of methods used in the studies to measure competencies is described, as well as their advantages and disadvantages. Secondly, valid and reliable methods found in the literature studies are presented. Finally, we report on the best methods per research purpose and per competency according to their advantages and disadvantages, and the validity and reliability of the measurement methods reported.

4.1. Type of methods

In the studies analysed, seven different measurement methods were found: questionnaires, rubrics, tests, observations, interviews, portfolios, and reflections. Questionnaires and rubrics are the most common (75%) assessment methods reported.

Questionnaires, which gather information from respondents through a set of written questions, were used in the form of self-assessment, where students assessed their own perceptions about their skills (Strauss and Terenzini 2005; Garcia Garcia et al. 2014) and attitudes (Douglas et al. 2014), or peer assessment, where students assessed each other (Zhang 2012). While questionnaires

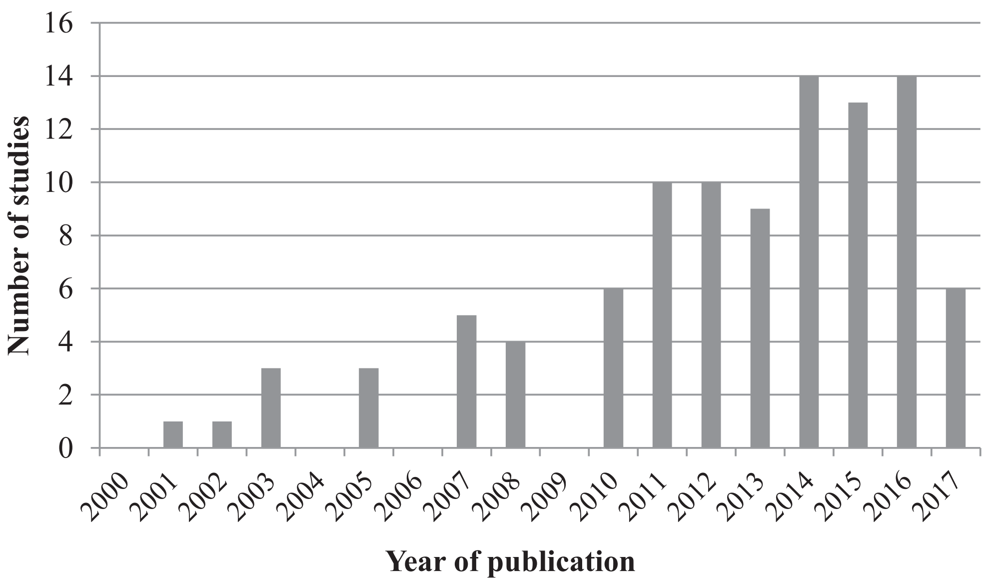

Figure 2. Number of published studies ($n = 99$) that studied communication, innovation/creativity, lifelong learning or teamwork competencies from 2000 to 2017 (October).

are easy to develop and require short time administration, questionnaires reported perceptions which are predisposed to bias (Douglas et al. 2014). Another issue observed was that the majority of the questionnaires used Likert scale questions and were performed at only one point in time, therefore ignoring the effect of social and process changes. To take into account this effect, some studies used pre- and post-questionnaires (Waychal 2014; Gerhart and Carpenter 2015; Ngaile, Wang, and Gau 2015), administered at the beginning and at the end of the programme or course, which allowed for observing changes in student competencies.

Rubrics, scoring methods with or without detailed descriptions of levels of performance, were used by faculty (Gerlick et al. 2011) or industry representatives (Hotaling et al. 2012) to assess written reports and oral presentations, designs projects, and capstone courses. Rubrics with detailed descriptions of levels of performance homogenised and guided the assessors (Flateby and Fehr 2008; Scharf 2014; Eichelman, Clark, and Bodnar 2015) because they increased inter-rater reliability and minimised subjectivity of the competency measurement process (Fila and Purzer 2012).

Tests, in the form of written and proof of concept tests, besides questionnaires and rubrics, were frequently used to measure innovation/creativity. Similar to questionnaires, they were administrated to measure skills or abilities, either after the course (Charyton, Jagacinski, and Merrill 2008; Charyton et al. 2011) or before and after the course (Shields 2007; Robbins and Kegley 2010). As with questionnaires, the use of pre- and post-test were considered good strategies to ensure the validity of the method (Cohen, Manion, and Morrison 2007).

Observations, which intended to observe student behaviour, were used as a stand-alone methodology to measure students behaviour by teaching assistants (Sheridan, Evans, and Reeve 2014) or peer-students (Pazos, Micari, and Light 2010), but also in combination with other methodologies, e.g. interviews (Dohaney et al. 2015). As a good practice, most of the observations were carried out using frameworks or rubrics to guide the measurement.

Interviews, in which an interviewer asks questions to an individual or group of interviewees, were also used as a stand-alone (Dolan et al. 2011), but mostly in combination with other instruments such as questionnaires (Barnes, Dyrenfurth, and Newton 2012; Dunai et al. 2015; Eichelman, Clark, and Bodnar 2015). Both observations and interviews are time-consuming for assessors and they require training, however in the case of observations they provide authentic student behaviour and attitudes, and interviews allow depth and flexibility of student responses. An alternative to common observations used by (Besterfield-Sacre et al. 2007) is work sampling observations. This type of observation takes place in floating-length intervals instead of full-time observation. This method, used to measure teamwork in four different learning environments, reported improvement in the cost-effectiveness of the observation method.

The least used methods were *portfolios* (Martínez-Mediano and Lord 2012; Wu, Huang, and Shadiev 2016) and *reflections* (Bursic, Shuman, and Besterfield-Sacre 2011). The portfolios consisted of a compilation of deliverables developed by students as part of their coursework, that shows meaningful learning. The data of portfolios were coded to demonstrate students' recognition of the need for and ability to engage in lifelong learning (Wu, Huang, and Shadiev 2016) and to measure the influence of a Moodle learning platform on students' creativity (Martínez-Mediano and Lord 2012).

Reflections included students reflecting on and describing their competency learning of a competency. Portfolios and reflections, as well as observations, were used to support the results obtained by other methods, such as tests, rubrics and questionnaires. We suspect that the low frequency found of these methods can likely be explained by the relatively large amount of time and work required by faculty members to use these instruments. The use of multiple methods was also reported in other studies present in the review. This is discussed in more detail in the next sections.

4.2. Validity and reliability

More than half of the methods presented in the 99 studies did not describe the theoretical background or research behind their metric designs. Only 39 studies (32 measurement methods) went beyond that and reported validity and reliability properties (Appendix E). Of these studies, 7 measured communication, 6 lifelong learning, 6 teamwork, and 9 measured innovation/creativity. Only 4 methods measured more than one competency: communication and innovation/creativity (Hernandez-Linares et al. 2015), communication and teamwork (Immekus et al. 2005; Fini and Mellat-Parast 2012), and communication, lifelong learning and teamwork (Strauss and Terenzini 2005).

On the one hand, in some studies a number of techniques were used to demonstrate validity: review of items or content from previous literature; review of experts and students' opinion about the content of the assessment; correlations between tests which intend to measure the same construct; use of control and experimental groups; confirmatory and factor analyses; and testing of the method as a pilot study. Reliability properties relied on internal consistency and inter-rater reliability. On the other hand, validity and reliability measurements were overlooked in other studies, i.e. they did not define the content being measured which immediately violated the definition of content validity.

It was also found that methods which presented reliable and valid measurements in previous studies were reused, such as Modified Strategies for Learning Questionnaire (Lord et al. 2011; Amelink et al. 2013), Abreaction Test for Evaluating Creativity (Clemente, Vieira, and Tschimmel 2016), Critical Thinking Assessment (Vila-Parrish et al. 2016), Index of Learning Styles (Waychal 2014), Torrance Test of Creativity Thinking (Shields 2007; Robbins and Kegley 2010; Wu, Huang, and Shadiev 2016), Lifelong Learning Scale (Kirby et al. 2010; Chen, Lord, and McGaughey 2013), and Self-Assessment of Problem Solving Strategies (Douglas et al. 2014). The convenience of using existing valid methods will be discussed later on.

4.2.1. Methods per assessment purpose

We intended to find out how the type of method could be related to the purpose of the measurement. This is important when creating or choosing a method because the design of a method may not be appropriate for a different purpose. For this reason, the distribution of the methods per measurement purpose was listed in Table 1 and the frequencies were analysed to verify what type of methods were more widespread per measurement purpose.

More than half of the studies reported on methods which were used to evaluate course and programme effectiveness to enhance the quality of teaching and student learning. The most frequent (63%) method used for this purpose was questionnaires. They were used to ask students about how the course prepares them for a competency (Baral et al. 2014; Gerhart and Carpenter 2015). Questionnaires alone, unless the sample size is large enough to have statistically significant results, are not a good practice, because they report self-perceptions which are subjective to bias. However, questionnaires used in combination with other methods such as portfolios (Martínez-Mediano and Lord 2012), interviews (Dunai et al. 2015) and observations (Blanco, López-Forniés, and Zarazaga-Soria 2017) showed that the courses stimulate the development of competencies in students. For example, in the study of Martínez-Mediano and Lord (Martínez-Mediano and Lord 2012), the use of portfolios confirmed the results of the questionnaire that the intervention had improved students' ability in lifelong learning. Similarly, a combination of interviews conducted by an external researcher and questionnaires given to students proved that the project-based learning promoted teamwork competencies (Dunai et al. 2015).

The second most frequent purpose (26%) was to assess students' performance with the purpose of giving summative grading and formative feedback. The former is used to provide student grades at the end of the curricular activity to certify students' achievements, and the latter is used to provide feedback to improve students' learning (Biggs 2003). Few studies (only 7% of the studies) which reported formative feedback were found. The results show that rubrics were the most frequent (62%) method used to grade students (Fila and Purzer 2012) and to provide formative feedback to students (Ahmed 2017). Rubrics were considered good practices for this type of measurement

Table 1. Distribution of methods with measurement purpose (1- to evaluate course and programme effectiveness to enhance the quality of teaching and the student learning experience, 2- to assess students' performance with the purposes of giving summative grading at the end of courses and/or providing formative feedback to students, and 3- to measure students' abilities in order to characterise students populations.) and competencies (CM - communication, LLL - lifelong learning, TW - teamwork, IC - innovation/creativity, and > C - more than one competency)

		Questionnaires	Rubrics	Tests	Observations	Interviews	Multiple methods
(1) Evaluate course and programme effectiveness	CM	4	5	-	-	-	3
	LLL	1	5	-	-	-	4
	TW	5	-	1	-	-	2
	IC	3	1	4	-	-	4
	>C	7	1	-	-	-	5
	Total	20	12	5	-	-	18
(2) Assess students' performance	CM	-	4	-	-	-	-
	LLL	1	3	-	-	-	-
	TW	2	1	-	3	1	1
	IC	-	2	-	-	-	-
	>C	-	6	-	-	-	2
	Total	3	16	-	3	1	3
(3) Measure student abilities	CM	1	-	1	-	-	1
	LLL	3	-	1	-	-	1
	TW	-	-	-	-	-	-
	IC	-	-	1	-	-	-
	>C	2	-	-	-	-	-
	Total	6	-	3	-	-	2
Combination of (1) and (2)	CM	-	1	-	-	-	3
	LLL	-	-	-	-	-	-
	TW	-	-	-	-	-	-
	IC	-	-	-	-	-	2
	>C	-	-	-	-	-	1
	Total	-	1	-	-	-	6
Total		29	29	8	3	1	29

purpose (Fila and Purzer 2012), for the reason that they were objective checklists based on student learning outcomes that allowed assessors to grade students, and to provide feedback.

The third form of measurement (11%) was aimed at measuring students' abilities in order to characterise student populations. More than half of these methods were questionnaires. For example, Strauss and Terenzini (Strauss and Terenzini 2005) aimed at assessing a large population of 4558 graduating seniors in seven engineering fields in more than one competency (e.g. communication, lifelong learning and teamwork) on a five-point Likert scale. Moreover, (Chen, Lord, and McGaughey 2013) conducted a cross-sectional study with 356 engineering student of five different fields and major. In this study, students were asked to evaluate their abilities for lifelong learning. Self-perception questionnaires were considered an acceptable strategy (Strauss and Terenzini 2005; Chen, Lord, and McGaughey 2013) when the aim was to evaluate a large population.

Within the three purposes (assess student learning, evaluate course/programme effectiveness and characterise student abilities), a limited number of studies used qualitative methods (e.g. observations, interviews, portfolios and reflections). This limitation will be addressed in the discussion.

4.3. Measurement methods per competency

A summary of the criteria found per competency is reported below, as well as a definition formulated for each competency based on the studies included. In addition, the best measurement methods per competency are suggested. This information may assist assessment developers in the development of their own competency assessment and evaluation schemes.

4.3.1. Competency definitions and measuring criteria

As stated by (Shuman, Besterfield-Sacre, and McGourty 2005), the lack of consensus on the definitions of the competencies creates difficulties in their measurement process. For this reason, we were interested to investigate how the studies define the competencies under study. A lack of competency definitions in the studies was found. Of all of them, only 17 studies explicitly define the competencies they were studying. Lack of definitions bias understanding when performing the measurement, and prejudice the replication of the studies. Since competency terms have various meanings depending on the context, it is problematic to assume that the competencies have the same synonym and do not warrant a definition.

For the studies that were not providing any definition for the competencies, we decided to investigate the criteria that were used to provide clarity and measure these competencies. Although 5% of the studies did not provide any criteria to establish the competencies, using only a Likert scale to rate the self-perceived level of the competencies undefined, such as in (Moalosi, Molokwane, and Mothibedi 2012), the analysis of the 99 studies disclosed several criteria used to measure the attainment levels in the four competencies. The criteria found for each competency, their definition and the corresponding studies are listed in Appendixes A, B, C and D, respectively. In the analysis of the results, we make no distinction on the purpose of the studies as our primary interest is to evaluate the criteria used to measure the attainment levels of competencies.

4.3.2. Communication (Appendix A)

Among the 44 studies that measured attainment levels in communication, 31 evaluated oral communication and 24 written communication. There were 16 studies that reported on both oral and written communication. Out of the 31 studies which looked at oral communication, 16 considered it as a single criterion without sub-division. The same was found for written communication (15 out of 24 studies).

A few studies which look at other communication criteria than oral and written communication were found. These criteria included *self-confidence* (4), *achieve/convey ideas* (3), *self-exposure* (2), *listening* (2), *reading* (1), and *client interaction* (1). These criteria suggest that communication for engineers is more than just oral and written communication (Wilkins, Bernstein, and Bekki 2015). It also involves listening actively, carrying general conversations, showing understanding by means of opinions or reactions on what is discussed, and self-exposure to conversations in order to interact with others and to create networking.

Based on the criteria listed above and the definitions found in studies such as (Immekus et al. 2005), (Wilkins, Bernstein, and Bekki 2015), we propose to use the following definition of communication: *communication* is 'the ability to show understanding and to carry technical/non-technical written/oral presentations and discussions depending on the audience where the feedback loop of giving and receiving opinions, advises and reactions is constant'.

To measure communication, valid methods (Appendix E) were found. (Eichelman, Clark, and Bodnar 2015) and (Galván-Sánchez et al. 2017) used rubrics to measure student performances in demonstrating written and oral communication, respectively. Also, (Frank et al. 2015) has objectively measured students' performance on written communication using two valid methods (the VALUE rubric the CLA+). (Wilkins, Bernstein, and Bekki 2015), on the other hand, validated a test that measures not only student self-perceived knowledge in communication skills (such as active listening, assertive self-expression, and receiving and responding to feedback), and their confidence to use these skills, but also their ability to apply these communication skills.

4.3.3. Lifelong learning (Appendix B)

The top five most frequently used criteria for lifelong learning competency were found to be *self-reflection* (17 studies), *locating and scrutinizing information* (16), *willingness, motivation and curiosity to learn* (11), *creating a learning plan* (10), and *self-monitoring* (6).

On the basis of the definitions present in the studies (Coşkun and Demirel 2010; Martínez-Mediano and Lord 2012) and the criteria found, we define *lifelong learning* as 'the intentional and active personal and professional learning that should take place in all stages of life, and in various contexts with the aim of improving knowledge, skills and attitudes'.

When it comes to reporting validity, one point in time self-assessment methods (Coşkun and Demirel 2010; Douglas et al. 2014) reported on validity measurements. On the other hand, EPSA (Ater Kranov et al. 2008; Ater Kranov et al. 2011; Ater Kranov et al. 2013; Schmeckpeper et al. 2014), another method that reports validity, goes beyond self-assessment and measures student performance on lifelong learning competencies during a specific task.

4.3.4. Teamwork (Appendix C)

For teamwork, criteria such as *interacting with others* (18 studies), *manage team responsibility* (15), *team relationship* (15), *communicating between group members/others* (9), and *contribution of ideas/ solutions/work* (9) were found to be the top 5 most frequently used criteria. Criteria such as *problem-solving and decision making* (8), and *encourage the group to contribute* (7) were also often named. Therefore, based on these criteria and the definitions present in the studies (Immekus et al. 2005; Valdes-Vasquez and Clevenger 2015), we define *teamwork* as 'an interactive process between a group of individuals who are interdependent and actively work together using their own knowledge and skills to achieve common purposes and outcomes which could not be achieved independently'.

The valid methods present in the review provide some adequate examples to measure teamwork. For example, rubrics were used to assess students' teamwork in capstone courses and the correlation between faculty and teaching assistant assessor was shown (Gerlick et al. 2011). In (Bringardner et al. 2016), both pre- and post-questionnaires were carried out to consider the effect of social and process changes in the measurement of student competency. Finally, (Besterfield-Sacre et al. 2007) provided a valid behavioural observation method which, however more time and resource consuming, proved that teamwork was accomplished.

4.3.5. Innovation/Creativity (Appendix D)

From the 24 studies which looked at innovation/creativity, 7 studies referred to innovation and 17 studies reported creativity. The low number of papers studying innovation may be an indication that only a small number of curriculum elements go beyond the design process and also focus on the idea or solution implementation step; as a consequence, measuring creativity levels is often deemed enough. Both innovation and creativity measurement criteria were found to focus mainly on *flexibility* (15 studies), *originality* (13), *fluency* (7), *elaboration* (7), *connection* (4), and *scaling information* (4).

On the basis of the criteria and definitions found in the studies (Fila and Purzer 2012; Amelink et al. 2013), we propose the following definition: *Innovation/Creativity* is 'the ability to generate ideas and move from their design to their implementation, thereby creating solutions, products and services for existing or future needs'.

For innovation/creativity, some valid methods were reused from previous studies. For instance, the Torrance Test of Creativity Thinking, that is validated in many studies (Shields 2007; Robbins and Kegley 2010; Wu, Huang, and Shadiev 2016), but requires trained assessors and is very costly. Other valid methods reported on are the Index of Learning Styles that measures innovation based on student preferences on a sensing/intuition scale (Waychal 2014), and the Modified Strategies for Learning Questionnaire that measures the perceptions of student learning behaviours in innovation skills (Amelink et al. 2013). More objective methods that measured student performance in demonstrating innovation rather than self-perceived are the Abreaction Test for Evaluating Creativity used in (Clemente, Vieira, and Tschimmel 2016) and the VALUE rubric used in (Vila-Parrish et al. 2016).

While analysing the criteria used in the studies, overlaps in the four competencies studied were found. This is not part of the scope of this review, so we will not go into detail. This finding confirms, however, the need to provide a definition for the competencies under study. As the underlying criteria depend on the definition, future studies should provide both competency definitions and underlying criteria so that conflicting elements can be avoided and coherent competency measurements carried out.

5. Discussion

The number of studies which looked at students' transversal competencies such as communication, innovation/creativity, lifelong learning, and teamwork competencies has grown over the last 17 years (Figure 2). This progression is likely indicative of the importance of these competencies for engineering students success in the labour market and the increase of their integration in engineering curricula (Passow and Passow 2017).

This systematic review on competency measurement shows that measuring competency levels has become extremely important to assess student performance in courses or programmes, to certify the level of courses and curricula, and to characterise student abilities. Based on the accuracy, validity and reliability of the methods analysed, the time and cost of their implementation, and their practicality for a specific purpose, we give recommendations to aid educators and researchers to further measure competencies, in terms of the best measurement methods, the importance of competency definitions and validity and reliability properties. Also, we offer principles to be applied in the creation of reliable and valid measurement methods.

5.1. Best measurement methods – for educators

To grade students and to provide feedback, we argue that it is not enough to ask students if they perceive competency improvements. However, the accuracy of self-assessment has been considered poor (Ward, Gruppen, and Regehr 2002). Methods that measure students demonstrating certain competencies would be more appropriate (Besterfield-Sacre et al. 2007). Rubrics can be used as a checklist to verify whether students demonstrate the pre-defined competencies and at which level (Fila and Purzer 2012). When rubrics are objectively created and validated to measure students' behaviours, they are great measurement methods that improve inter-marker consistency and reduce marker bias effects (Flateby and Fehr 2008; Scharf 2014; Eichelman, Clark, and Bodnar 2015). In addition, this consistency can be optimised with the use of more than one rater or grader and the standardisation of the scales according to graders' scores (Ward, Gruppen, and Regehr 2002). These techniques were proposed in (Ward, Gruppen, and Regehr 2002) as alternatives to reduce the issues of the efficacy of self-assessment. Rubrics can also be used for large samples, as experienced by the second author of this review (Saunders-Smits and Melkert 2011). Moreover, rubrics are useful not only to conduct summative assessment but also to provide individual feedback to strengthen detected points in students that need improvement. However, this form of assessment was little addressed by the studies reviewed.

Alternative measurement methods that are adequate to measure student behaviour are observations. However, they are very time and resource consuming. To reduce these issues, work sampling observation as validated in (Besterfield-Sacre et al. 2007) can be a very valuable method, because it reduces the amount of observation time necessary to assess students behaviour and consequently it is less labour intensive and time-consuming. We consider those behavioural measurements when based on clear criteria effective tools to provide summative and formative feedback. In Table 2, a set of practical guidelines for implementation in education is listed.

5.2. Best measurement methods – for researchers

For researchers who are willing to measure student competencies to evaluate courses or programmes or simply to characterise a student population, we argue that questionnaires and tests that measure perceptions are considered adequate methods for these purposes when limited time and resources are available and large samples are present. Self-report methods can be easily developed and administered, and when favourable validity and reliability properties are present, meaningful inferences can be drawn from the data analysis (Immekus et al. 2005). However, when using these methods (questionnaires or tests), we recommend the use of time triangulation by employing pre-

Table 2. Guidelines for reliable and valid measurement methods.

For educators	For researchers
1) Define each competency.	
2) Create sub-components of competencies and defined & describe each dimension or level of mastery.	
3) When grading or giving feedback, use checklists with the pre-defined sub-components and levels of competencies (Fila and Purzer 2012).	3) When measuring learning or growth, measure student performance on a competency before and after instruction (Cohen, Manion, and Morrison 2007) or ask students for extremes: what they learn the most and the least (Ward, Gruppen, and Regehr 2002).
4) Standardise scales/checklists i.e. create familiarity with the levels/dimensions of the scales and rescale them based on graders' assessment scores (Ward, Gruppen, and Regehr 2002).	4) Analyse the reliability and validity properties of the measurement to evaluate both the accuracy of the method and whether the method measures what it intends to measure (Cohen, Manion, and Morrison 2007).
5) Use more than one grader (Ward, Gruppen, and Regehr 2002).	5) Use multiple methods when corroboration, elaboration, clarification and expansion of the results is needed (Creswell and Clark 2007).
6) Analyse the level of agreement between the graders testing inter-rater reliability (Cohen, Manion, and Morrison 2007).	
7) When using self or peer assessment questionnaires, ask students for aspects that they need the most and least improvement (Ward, Gruppen, and Regehr 2002).	

and post-questionnaires (Waychal 2014; Gerhart and Carpenter 2015; Ngaile, Wang, and Gau 2015) or pre- and post-tests (Shields 2007; Robbins and Kegley 2010) to rectify the omission of social changes and processes caused by one-time assessment (Cohen, Manion, and Morrison 2007). Instead, self-assessment can be done by ranking competencies where students have to identify their own strengths and weaknesses, which are the extremes of the scales (Ward, Gruppen, and Regehr 2002). This method was not used in any study of this review but we recommend it, because it increases the accuracy of judging one's own performance which has been a great concern in literature (Ward, Gruppen, and Regehr 2002; Eva and Regehr 2005). Also, it is considered ideal to self-directed students' learning and to give formative feedback (Ward, Gruppen, and Regehr 2002; Eva and Regehr 2005).

Another strategy to increase validity in the case of self-perceptions is the use of multiple methods to measure the full umbrella of criteria of one or more competencies. The advantage of this is that combining different methods yield the most comprehensive information from different perspectives and a more complete understanding of the research problem (Creswell and Clark 2007). Studies in this review (Barnes, Dyrenfurth, and Newton 2012; Amelink et al. 2013; Eichelman, Clark, and Bodnar 2015) suggested that the content validity of the results of the assessment increased because the results from different methods could be compared, explained and verified, and the strengths and weaknesses of the methods could be drawn and minimised, respectively. For example, the use of rubrics alongside interviews benefit from their individual power: the rubric with described levels guides the assessor and reduces inconsistencies in the assessment because the measurement criteria are clear, delimited and objective, and the interviews offer more comprehensive information about students' competency development and since interviews are more flexible richer details can be obtained (Eichelman, Clark, and Bodnar 2015). Alternatively, researchers could employ a combination of questionnaires, which are straightforward and require little administration, with observations, which provide in situ data from the situations which are taking place (Amelink et al. 2013). Guidelines for researchers to create reliable and valid measurement methods are listed in Table 2.

At the moment, works published on competency measurements present in literature tend to rely heavily on the course evaluation only, and we were unable to find any longitudinal studies where students were followed in their years after completion of those courses or even after graduation. In future, educational researchers could consider using, if ethical boards allow, and willing participants are found, e.g. portfolios or interviews to perform longitudinal studies by collecting data from the same group of students at different points in their life, thus following the level of competency improvement of the students during their time at their institution and ideally also after graduation in their working life.

5.3. Importance of definitions and validity

We observe that in some studies there is an effort in developing valid competency measurements. Some described competencies based on literature, industry and students feedback; Others used multiple methods to improve content validity or conduct factor analyses to increase construct validity. In addition, some studies used existing validated measurement methods. Choosing existing valid and reliable instruments may form a helpful option for assessment developers and instructors to measure competencies in students. However, learning outcomes, competencies and course or programme settings should be carefully considered and compared to the conditions of the existing studies, to ensure their applicability. Re-evaluation of validity and reliability are still necessary when implemented in a new situation (Cohen, Manion, and Morrison 2007).

Although robust methods were found, some studies did not define the content being measured and therefore they overlook content validity. Lack of consensus on the definition of the transversal competencies was a cause of difficulties in the process of competency measurement (Shuman, Besterfield-Sacre, and McGourty 2005). Likewise, the lack of definitions may hinder the measurement of competencies. In this literature review, 83% of the studies identified and included did not present a definition of the assessed transversal competencies, and 5% did not provide any criteria to establish the competencies. What were the perceived definitions of students or instructors when using these methods without definitions or descriptions? It is possible and acceptable that the definitions of competencies determined by different entities could be different. However, it should be clear for all involved parties what the definitions of the terms used are. Only with clear definitions and descriptions can measurement of competency attainment levels be understandable and valuable.

Overall, competency level measurement would benefit from better method design and validity evidence. The only way to ensure that the results obtained from the competency measurements are accurate and can be properly interpreted is through a clear and described assessment design and by carrying out validity and reliability measurements. Only 39 studies had methods that consistently measured reliability and validity. This means that accurate results can be extracted from only 39 out of the 99 studies. Validity and reliability measurements provide feedback to both researchers and educators whether methods measure the initial proposed concept and allow them to engage in subsequent revision and improvement of the measurement methods.

6. Conclusion, limitations and recommendations

This systematic review set out to inform the creation of reliable and valid measures of student mastery of competencies in communication, lifelong learning, innovation/creativity and teamwork in engineering education. We analysed measurement methods of 99 studies published in the last 17 years. This review described the type of methods that measure the four previously mentioned competencies, and their advantages and disadvantages, and validity and reliability properties based on the studies analysed. From the analysis of these findings, the best methods per purpose and competency are presented. Additionally, a definition for each competency and its underlying criteria are reported to assist assessment developers in the design of their own competency assessment and evaluation schemes.

Some limitations in the current studies that measure competencies have arisen regarding competency definitions and validity and reliability measurements. The analysis showed that a large number of studies lack a clear definition of the selected competency. Based on these issues, we shed a light on the importance of providing clear definitions and underlying criteria for the competencies under study. As such, we created a clear definition for each competency.

Moreover, less than half of the studies presented evidence of validity and reliability measurements. This result shows that a clear need to set professional standards when measuring competencies are needed and that future studies should report on validity and reliability measurements.

Questionnaires and rubrics were the methods mostly used to measure these competencies. We argue that both are adequate methods when properly validated with the techniques present in this review. Questionnaires, applied in the form of pre and post-questionnaires, are particularly useful for assessors/researchers to evaluate course or programme effectiveness and characterise students' abilities in the presence of large student populations. This review also showed the usefulness of combining methods (particularly questionnaires with interviews or observations) to increase the validity of the studies. As such, researchers are encouraged to use multiple methods when evaluating the effectiveness of courses or programmes to stimulate student competencies, and when characterising students' abilities.

On the other hand, rubrics benefit evaluators in the grading and feedback processes both for small or large populations when their scales are clearly defined according to course learning outcomes. Questionnaires that ask students for aspects that they need the most and least improvement are also a good practice. Alternatives are observations, portfolios and reflections, however they are labour intense and more time-consuming.

While there is a global concern and effort in engineering education to measure competencies in communication, teamwork, lifelong learning and innovation shown in this review, engineering educators and future researchers should double their efforts to provide competency definitions and validate their measurement methods. We believe that time, energy and cost are undesirable limiting factors, but other issues such as lack of expertise and accuracy in the design and implementation of the measurement tool must be overcome. It may be worth as to why only a few studies provide explicit transversal competency measurement instruments. This may help improve the field of engineering education in the area of competency measurement.

A potential limitation of this systematic review is that powerful papers might have been left out because we might have excluded alternative terms used to name the four competencies. The review was also limited to engineering students, three databases and the past 17 years. It may be worthwhile in the future endeavours to expand the review to the fields of science, technology and mathematics, other databases and possibly look at papers before 2000.

The PREFER project will use the lessons learned in this review to create a measurement tool that measures students' mastery levels in courses. A valid and reliable method will be designed. We will define competencies and their subcomponents to address the full extent of each competency; perform confirmatory and exploratory factor analyses and test the internal structure of the measurement scales. The outcomes of the tool will be triangulated using student reflections as this is the most feasible method within the scope of our project.

Acknowledgments

This work was supported by Erasmus+ programme of the European Union (grant agreement 575778-EPP-1-2016-1-BE-EPPKA2-KA) and is part of the PREFER project (http://www.preferproject.eu/).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Erasmus+ programme of the European Union: [Grant Number Grant agreement 575778-EPP-1-2016-1-BE-EPPKA2-KA].

Notes on contributors

Mariana Leandro Cruz received the BSc and MSc degrees in Biomedical Engineering from the Instituto Superior Técnico, University of Lisbon. She is currently developing the PhD in engineering education at Delft University of Technology,

Faculty of Aerospace Engineering in the Netherlands. Her research interest include engineering education, competencies, competency measurement, and course development.

Dr. ***Gillian N. Saunders-Smits*** is a Senior Lecturer/Associate Professor at Aerospace Engineering faculty in Delft University of Technology. She has a Master Degree in Aerospace Engineering from Delft University of Technology, Faculty of Aerospace Engineering in the Netherlands and a PhD in Aerospace Engineering Education from the same institute. She has extensive teaching experience as well as a broad experience in curriculum development (BSc and MSc) and educational leadership. She has served as the Faculty's online education coordinator and actively develops online courses and platforms for open learning. In 2017, she has been appointed board member of the Steering Committee of the Société Européenne pour la Formation des Ingénieurs (SEFI).

Prof. Dr. ***Pim Groen*** is the professor of SMART Materials at Aerospace Engineering faculty in Delft University of Technology and the Programme Manager of Holst Centre, TNO, the Netherlands. He studied chemistry and obtained the PhD in Materials Science from the University of Leiden in the Netherlands. He has extensive experience as a scientist and from 2012 he has been appointed as a part-time professor where he can spread his scientific knowledge and industry experience with Aerospace Engineering students at Delft University of Technology.

References

Ahmed, S. 2017. "*Teaching and Assessing Lifelong Learning in Laboratory Courses*." Proceedings of the Canadian engineering education Association (CEEA17), Toronto, Ontario.

Amante, B., A. Lacayo, M. Pique, S. Oliver, P. Ponsa, and R. Vilanova. 2010. "*Evaluation of Methodology PBL Done by Students*." IEEE Transforming engineering education: creating Interdisciplinary skills for complex global environments, Dublin, Ireland, pp. 1-21.

Amelink, C. T., B. A. Watford, G. R. Scales, M. J. Mohammadi-Aragh, and D. Farquhar-Caddell. 2013. "*Defining and Measuring Innovative Thinking Among Engineering Undergraduates*." 120th ASEE Annual conference & exposition, Atlanta, Georgia.

American Society for Engineering Education. 1994. *Engineering Education for a Changing World: A Joint Project by the Engineering Deans Council and Corporate Roundtable of the American Society for Engineering Education*. Washington, DC: American Society for Engineering Education.

Ater Kranov, A., C. Hauser, O. Robert, and L. Girardeau. 2008. "*A Direct Method for Teaching and Assessing Professional Skills in Engineering Programs*." 115th ASEE Annual conference & exposition, Pittsburgh, Pennsylvania.

Ater Kranov, A., R. L. Williams, P. D. Pedro, E. R. Schmeckpeper, S. W. Beyerlein, and J. McCormack. 2013. "*A Direct Method for Teaching and Measuring Engineering Professional Skills for Global Workplace Competency: Overview of Progress of a Current NSF-Sponsored Validity Study*." Proceedings of the 2013 ASEE International Forum, Atlanta, Georgia.

Ater Kranov, A., M. Zhang, S. W. Beyerlein, J. McCormack, P. D. Pedro, and E. R. Schmeckpeper. 2011. "*A Direct Method for Teaching and Measuring Engineering Professional Skills: a Validity Study*." 118th ASEE Annual conference & exposition, Vancouver, Canada.

Backer, P. R. 2016. "*Assessment of a Writing Workshop Model for First-Year Engineering Students*." 123rd ASEE Annual conference & exposition, New Orleans, Louisiana.

Badcock, P. B. T., P. E. Pattison, and K. Harris. 2010. "Developing Generic Skills Through University Study: a Study of Arts, Science and Engineering in Australia." *Higher Education* 60 (4): 441–458.

Badran, I. 2007. "Enhancing Creativity and Innovation in Engineering Education." *European Journal of Engineering Education* 32 (5): 573–585.

Baral, L. M., R. Muhammad, C. V. Kifor, and I. Bondrea. 2014. "*Evaluating the Effectiveness of Problem-Based Learning (PBL) Implemented in the Textile Engineering Course-a Case Study on Ahsanullah University of Science and Technology, Bangladesh*." 6th Balkan Region conference on engineering and Business education & 5th International conference on engineering and Business education & 4th International conference on innovation and Entrepreneurship, pp. 111-114.

Barnes, S. K., M. J. Dyrenfurth, and K. A. Newton. 2012. "*Mixed Methods Model Approach for Measuring International Engineering, Design and Technology Student Exchange Programs*." 119th ASEE Annual conference & exposition,, San Antonio, Texas.

Barr, N. 2016. "*Extending WID to Train Mechanical Engineering GTAs to Evaluate Student Writing*." 123rd ASEE Annual conference & exposition, New Orleans, Los Angeles.

Bartram, D. 2005. "The Great Eight Competencies: a Criterion-Centric Approach to Validation." *Journal of Applied Psychology* 90 (6): 1185.

Bennett, N., E. Dunne, and C. Carré. 2000. *Skills Development in Higher Education and Employment*. Buckingham, United Kingdom: The Society for Research into Higher Education and Open University Press.

Besterfield-Sacre, M., L. J. Shuman, H. Wolfe, R. M. Clark, and P. Yildirim. 2007. "Development of a Work Sampling Methodology for Behavioral Observations: Application to Teamwork." *Journal of Engineering Education* 96 (4): 347–357.

Biggs, J. 2003. *Teaching for Quality Learning at University: What the Student Does*. Buckingham, United Kingdom: The Society for Research into Higher Education & Open University Press.

Blanco, T., I. López-Forniés, and F. J. Zarazaga-Soria. 2017. "Deconstructing the Tower of Babel: a Design Method to Improve Empathy and Teamwork Competences of Informatics Students." *International Journal of Technology and Design Education* 27 (2): 307–328.

Bousaba, N. A., J. M. Conrad, J. L. Coco, M. Miri, and R. W. Cox. 2014. "*Incorporating Oral Presentations Into Electrical and Computer Engineering Design Courses: A Four-Course Study*." 121th ASEE Annual conference & exposition, Indianapolis, Indiana.

Brake, N. A., and J. Curry. 2016. "*The Impact of One-Credit Introductory Engineering Courses on Engineering Self-Efficacy: Seminar vs. Project-Based*." 123rd ASEE Annual conference & exposition, New Orleans, Louisiana.

Briedis, D. 2002. "Developing Effective Assessment of Student Professional Outcomes." *International Journal of Engineering Education* 18 (2): 208–216.

Bringardner, J., C. Leslie, G. W. Georgi, and A. M. D'Apice. 2016. "*Improving Efficacy in Group Projects with Teamwork Agreements*." 123rd ASEE Annual conference and exposition, New Orleans, Louisiana.

Bursic, K. M., L. J. Shuman, and M. Besterfield-Sacre. 2011. "*Improving Student Attainment of ABET Outcomes Using Model-Eliciting Activities (MEAS)*." 118th ASEE Annual conference & exposition.

Care, E., and R. Luo. 2016. "Assessment of Transversal Competencies: Policy and Practice in the Asia-Pacific Region." In *United Nations Educational, Scientific and Cultural Organization*, 307–366. Paris, France: UNESCO.

Charyton, C., R. J. Jagacinski, and J. A. Merrill. 2008. "CEDA: A Research Instrument for Creative Engineering Design Assessment." *Psychology of Aesthetics, Creativity, and the Arts* 2 (3): 147.

Charyton, C., R. J. Jagacinski, J. A. Merrill, W. Clifton, and S. DeDios. 2011. "Assessing Creativity Specific to Engineering with the Revised Creative Engineering Design Assessment." *Journal of Engineering Education* 100 (4): 778–799.

Chen, J. C., S. M. Lord, and K. J. McGaughey. 2013. "*Engineering Students' Development as Lifelong Learners*." 120th ASEE Annual conference & exposition, Atlanta, Georgia.

Clemente, V., R. Vieira, and K. Tschimmel. 2016. "*A Learning Toolkit to Promote Creative and Critical Thinking in Product Design and Development Through Design Thinking*." 2nd International conference of the Portuguese Society for engineering education (CISPEE), Vila-Real, Portugal, pp. 1-6.

Cohen, L., L. Manion, and K. Morrison. 2007. *Research Methods in Education*. Abingdon, Oxon: Routledge.

Colsa, Uruburu, Á I, Ortiz Marcos, and J. R. Cobo Benita. 2015. "Improving Engineering Students' Communication Competence: Designing Innovative Learning Strategies." *International Journal of Engineering Education* 31 (1 (B)): 361–367.

Coşkun, Y. D., and M. Demirel. 2010. "Lifelong Learning Tendency Scale: the Study of Validity and Reliability." *Procedia-Social and Behavioral Sciences* 5: 2343–2350.

Council, N. R. 2013. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: National Academies Press.

Craps, S., M. Pinxten, H. Knipprath, and G. Langie. 2018. "*Wanted: Super (wo) man A Study to Define Professional Roles for Future Engineers by Distinctive Professional Competences*." Proceedings of the 46th SEFI Annual conference 2018, Copenhagen, Denmark.

Crawley, E., J. Malmqvist, S. Ostlund, and D. Brodeur. 2007. "Rethinking Engineering Education." *The CDIO Approach* 302: 60–62.

Creswell, J. W., and V. L. P. Clark. 2007. *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

Cropley, D. H. 2015. "Promoting Creativity and Innovation in Engineering Education." *Psychology of Aesthetics, Creativity, and the Arts* 9 (2): 161.

de-Juan, A., A. Fernandez del Rincon, M. Iglesias, P. Garcia, A. Diez-Ibarbia, and F. Viadero. 2016. "Enhancement of Mechanical Engineering Degree Through Student Design Competition as Added Value. Considerations and Viability." *Journal of Engineering Design* 27 (8): 568–589.

Dohaney, J., E. Brogt, B. Kennedy, T. M. Wilson, and J. M. Lindsay. 2015. "Training in Crisis Communication and Volcanic Eruption Forecasting: Design and Evaluation of an Authentic Role-Play Simulation." *Journal of Applied Volcanology* 4 (1): 12.

Dolan, D., M. Batchelder, J. McReynolds, K. Osberg, R. Koontz, P. Mahon, M. Keegan, and J. Weiss. 2011. "*Student Development in the co-Curriculum Through Values-Based Teaming*." 41st ASEE/IEEE Frontiers in education conference (FIE), Rapid City, South Dakota, pp. T3C-1-T3C-6.

Douglas, K. A., R. E. H. Wertz, M. Fosmire, Ş Purzer, and A. S. van Epps. 2014. "*First Year and Junior Engineering Students' Self-Assessment of Information Literacy Skills*." Proceedings of the 121st ASEE Annual conference & exposition, Indianapolis, Indiana.

Dunai, L. D., A. Prieto, M. Chillarón, and J. A. Antonino-Daviu. 2015. "*Education in Electric and Electronic Engineering via Students Involvement in Innovative Projects*." 41st Annual conference of the IEEE Industrial Electronics Society Yokohama, Japan, pp. 001862-001866.

Eichelman, K. M., R. M. Clark, and C. A. Bodnar. 2015. "*Assessing the Impact of Game Based Pedagogy on the Development of Communication Skills for Engineers*." 122th ASEE Annual conference & exposition, Seattle, Washington.

Engineering Accreditation Commission. 2000. *Criteria for Accrediting Engineering Programs*. Baltimore: ABET, inc.
Enszer, J. A., and M. Castellanos. 2013. "*A Comparison of Peer Evaluation Methods in Capstone Design*." 120th ASEE Annual conference & exposition, Atlanta, Georgia.
Eppes, T. A., I. Milanovic, and F. Sweitzer. 2012. "Towards Liberal Education Assessment In Engineering And Technology Programs." *Journal of College Teaching & Learning (Online)* 9 (3): 171.
Eva, K. W., and G. Regehr. 2005. "Self-assessment in the Health Professions: a Reformulation and Research Agenda." *Academic Medicine* 80 (10): S46–S54.
Fagerholm, F., and A. Vihavainen. 2013. "*Peer Assessment in Experiential Learning Assessing Tacit and Explicit Skills in Agile Software Engineering Capstone Projects*." 2013 IEEE Frontiers in education conference, Oklahoma City, Oklahoma, pp. 1723-1729.
Fila, N. D., and S. Purzer. 2012. "*Do Students Dream Beyond LEDs? Innovative Qualities of Student Solutions to an Idea Generation Task*." 199th ASEE Annual conference & exposition, San Antonio, Texas.
Fini, E., and M. Mellat-Parast. 2012. "Empirical Analysis of Effect of Project-Based Learning on Student Learning in Transportation Engineering." *Transportation Research Record: Journal of the Transportation Research Board* 2285: 167–172.
Flateby, T., and R. Fehr. 2008. "Assessing and Improving Writing in the Engineering Curriculum." *International Journal of Engineering Education* 24 (5): 901.
Frank, B. M., J. A. Kaupp, N. Simper, and J. Scott. 2015. "*Multi-method Longitudinal Assessment of Transferrable Intellectual Learning Outcomes*." 122nd ASEE Annual conference & exposition, Seattle, Washington.
Fries, R., B. Cross, J. Zhou, and C. Verbais. 2017. "How Student Written Communication Skills Benefit During Participation in an Industry-Sponsored Civil Engineering Capstone Course." *Advances in Engineering Education* 6 (1): 1–22.
Galván-Sánchez, I., D. Verano-Tacoronte, S. M. González-Betancor, M. Fernández-Monroy, and A. Bolívar-Cruz. 2017. "Assessing Oral Presentation Skills in Electrical Engineering: Developing a Valid and Reliable Rubric." *International Journal of Electrical Engineering Education* 54 (1): 17–34.
García, A. I., J. M. Fuentes, Á Ramírez-Gómez, P. Aguado, and F. Ayuga. 2015. "*Divulgative Technical Videos by University Students as a Process for the Acquisition of Cross-Curricular Skills in the Context of EHEA Degrees*." Proceedings of INTED2015 conference, Madrid, Spain.
Garcia, I. A., J. A. Calvo-Manzano, C. L. Pacheco, and C. A. Perez. 2015. "Software Engineering Education for a Graduate Course: A web-Based Tool for Conducting Process Improvement Initiatives with Local Industry Collaboration." *Computer Applications in Engineering Education* 23 (1): 117–136.
Garcia Garcia, M. J., C. Gonzalez Garcia, L. J. Fernandez Gutierrez Del Alamo, J. L. Casado Sanchez, and L. M. Martinez Muneta. 2014. "Assessing Creativity In Engineering Students: A Comparative Between Degrees and Students In First And Last Year." *International Journal of Engineering Education* 31 (1 (B)): 1–11.
Gerhart, A. L., and D. D. Carpenter. 2015. "*Creativity, Innovation, and Ingenuity Summer Enrichment Program—Collaborating with a Cultural Institution and Preliminary Assessment*." 122nd Annual conference & exposition, Seatle, Washington.
Gerlick, R., D. C. Davis, M. S. Trevisan, and S. A. Brown. 2011. "*Establishing Inter-Rater Agreement for TIDEE's Teamwork and Professional Development Assessments*." 118th ASEE Annual conference & exposition, Vancouver, British Columbia.
Heinis, T. B., I. Goller, and M. Meboldt. 2016. "*Multilevel Design Education for Innovation Competencies*." 26th CIRP design conference.
Hernandez-Linares, R., J. E. Agudo, M. Rico, and H. Sánchez. 2015. "Transversal Competences of University Students of Engineering." *Croatian Journal of Education: Hrvatski Časopis za Odgoj i Obrazovanje* 17 (2): 383–409.
Hernández, J. T., and C. Ramírez. 2008. "*Innovation and Teamwork Training in Undergraduate Computing Engineering Education*." Proceedings of 36th SEFI Annual conference, Aalborg, Denmark.
Hotaling, N., B. B. Fasse, L. F. Bost, C. D. Hermann, and C. R. Forest. 2012. "A Quantitative Analysis of the Effects of a Multidisciplinary Engineering Capstone Design Course." *Journal of Engineering Education* 101 (4): 630–656.
Huyck, M., D. Ferguson, and R. Wasserman. 2007. "*Assessing Factors Contributing to Undergraduate Multidisciplinary Project Team Effectiveness*." 2007 ASEE Annual conference & exposition, Honolulu, Hawaii.
Immekus, J. C., S. J. Maller, S. Tracy, and W. C. Oakes. 2005. "*Evaluating the Outcomes of a Service-Learning Based Course in an Engineering Education Program: Preliminary Results of the Assessment of the Engineering Projects in Community Service—EPICS*." 2005 ASEE Annual conference & exposition,, Portland, Oregon.
Jones, D. K., and M. Abdallah. 2013. "*Assessment of Communication and Teamwork Skills in Engineering Technology Programs*." Proceedings of 120th ASEE Annual conference & exposition, Atlanta, Georgia.
Kamp, A. 2016. *Engineering Education in the Rapidly Changing World*. Delft, the Netherlands: 4TU.Centre for Engineering Education.
Kaul, S., and R. D. Adams. 2014. "*Learning Outcomes of Introductory Engineering Courses: Student Perceptions*." 121st ASEE Annual conference & exposition, Indianapolis, Indiana.
Keh, L. K., Z. Ismail, and Y. M. Yusof. 2017. "Creativity among Geomatical Engineering Students." *International Education Studies* 10 (4): 43.
Kemp, I. J., and L. Seagraves. 1995. "Transferable Skills—can Higher Education Deliver?" *Studies in Higher Education* 20 (3): 315–328.

Keshavarz, M., and S. A. Vaziri. 2012. "Assessment of Student Professional Outcomes for Continuous Improvement." *International Journal of Innovation in Education* 1 (4): 324–334.
Kirby, J. R., C. Knapper, P. Lamon, and W. J. Egnatoff. 2010. "Development of a Scale to Measure Lifelong Learning." *International Journal of Lifelong Education* 29 (3): 291–302.
Knobbs, C. G., and D. J. Grayson. 2012. "An Approach to Developing Independent Learning and non-Technical Skills Amongst Final Year Mining Engineering Students." *European Journal of Engineering Education* 37 (3): 307–320.
Litzinger, T., J. Wise, S. Lee, T. Simpson, and S. Joshi. 2001. "*Assessing Readiness for Lifelong Learning*." 2001 ASEE Annual conference & exposition, Albuquerque, New Mexico.
Lopes, D. C., M. C. Gerolamo, Z. A. P. Del Prette, M. A. Musetti, and A. Del Prette. 2015. "Social Skills: A key Factor for Engineering Students to Develop Interpersonal Skills." *International Journal of Engineering Education* 31 (1): 405–413.
Lopez-Malo, A., S. Husted, J. V. Gutierrez, and N. Ramirez-Corona. 2016. "*Food Engineering Students' Creative Experiences in a Capstone Course*." 123rd ASEE Annual conference & exposition, New Orleans, Louisiana.
Lord, S. M., C. Stefanou, M. J. Prince, J. Chen, and J. D. Stolk. 2011. "*Student Lifelong Learning Outcomes for Different Learning Environments*." 118th ASEE Annual conference & exposition, Vancouver, B.C., Canada.
Markes, I. 2006. "A Review of Literature on Employability Skill Needs in Engineering." *European Journal of Engineering Education* 31 (6): 637–650.
Martínez-Mediano, C., and S. M. Lord. 2012. "Lifelong Learning Competencies Program for Engineers." *International Journal of Engineering Education* 28 (1): 130.
McMasters, J., and L. Matsch. 1996. "*Desired Attributes of an Engineering Graduate-An Industry Perspective*." Advanced measurement and Ground testing conference, New Orleans, Louisiana, pp. 2241.
Moalosi, R., S. Molokwane, and G. Mothibedi. 2012. "Using a Design-Orientated Project to Attain Graduate Attributes." *Design and Technology Education* 17 (1): 30–43.
Mourtos, N. J. 2003. "*Defining, Teaching and Assessing Lifelong Learning Skills*." 33rd ASEE/IEEE Frontiers in education conference, Boulder, Colorado.
Mourtos, N. J. 2014. "*Integrating General Education Outcomes Into a Senior Design Capstone Experience*." 2014 International conference on interactive Collaborative learning (ICL), Dubai, United Arab Emirates.
Narayanan, M. 2013. "*Assessment of Learning Based on the Principles of Discovery and Metacognition*." 120th ASEE Annual conference & exposition,, Atlanta, Georgia.
Ngaile, G., J. Wang, and J. Gau. 2015. "Challenges in Teaching Modern Manufacturing Technologies." *European Journal of Engineering Education* 40 (4): 432–449.
Organisation for Economic Co-Operation Development. 2005. *The Definition and Selection of key Competencies: Executive Summary*. Paris, France: OECD.
Ortiz, M. I., L. I. Ballesteros Sánchez, T. Prieto Remón, and A. Uruburu Colsa. 2015. "*Strengthening and Measuring Project Management Competences of Engineering Students*." Proceedings of INTED2015 conference, Madrid, Spain.
Passow, H. J., and C. H. Passow. 2017. "What Competencies Should Undergraduate Engineering Programs Emphasize? A Systematic Review." *Journal of Engineering Education* 106 (3): 475–526.
Pazos, P., N. Magpili, Z. Zhou, and L. J. Rodriguez. 2016. "*Developing Critical Collaboration Skills in Engineering Students: Results From an Empirical Study*." 123rd ASEE Annual conference & exposition, New Orleans, Louisiana.
Pazos, P., M. Micari, and G. Light. 2010. "Developing an Instrument to Characterise Peer-led Groups in Collaborative Learning Environments: Assessing Problem-Solving Approach and Group Interaction." *Assessment & Evaluation in Higher Education* 35 (2): 191–208.
Petticrew, M., and H. Roberts. 2006. *Systematic Reviews in the Social Sciences: A Practical Guide*. Oxford, England: Blackwell.
Ramirez, C., N. A. Jimenez, and J. T. Hernandez. 2007. "*Teamwork and Innovation Competencies: A First Semester Engineering Students Hands-on Course*." International conference of engineering education (ICEE), Coimbra, Portugal.
Ramírez-Echeverry, J. J., F. A. Olarte Dussán, and A. García-Carillo. 2016. "Effects of an Educational Intervention on the Technical Writing Competence of Engineering Students." *Ingeniería e Investigación* 36 (3): 39–49.
Robbins, T. L., and K. Kegley. 2010. "Playing with Thinkertoys to Build Creative Abilities Through Online Instruction." *Thinking Skills and Creativity* 5 (1): 40–48.
Ryan, R. G. 2012. "*Assessment of a New Design Stem Course Sequence*." 119th ASEE Annual conference & exposition, San Antonio, Texas.
Saorín, J. L., D. Melian-Díaz, A. Bonnet, C. C. Carrera, C. Meier, and J. De La Torre-Cantero. 2017. "Makerspace Teaching-Learning Environment to Enhance Creative Competence in Engineering Students." *Thinking Skills and Creativity* 23: 188–198.
Saunders, K., C. E. Glatz, M. Huba, M. Griffin, S. Mallapragad, and J. V. Shanks. 2003. "*Using Rubrics to Facilitate Students' Development of Problem Solving Skills*." 2003 ASEE Annual conference & exposition.
Saunders-Smits, G. N., and J. Melkert. 2011. "*Innovative Grading for Design Exercises - A Case Study From Aerospace Engineering*." International Association of Societies of design research (IASDR) - 4th World conference on design research, Delft, the Netherlands.
Scharf, D. 2014. "*Instruction and Assessment of Information Literacy among STEM Majors*." 4th IEEE Integrated STEM education conference (ISEC), Princeton, New Jersey, pp. 1–7.

Schmeckpeper, E. R., A. Ater Kranov, S. W. Beyerlein, J. McCormack, and P. D. Pedro. 2014. "*Using the EPSA Rubric to Evaluate Student Work in a Senior Level Professional Issues Course*." 121st ASEE Annual conference & exposition, Indianapolis, Indiana.

Schoepp, K., M. Danaher, and A. Ater Kranov. 2016. "*The Computing Professional Skills Assessment: An Innovative Method for Assessing ABET's Student Outcomes*." 2016 IEEE global engineering education conference (EDUCON), Abu Dhabi, United Arab Emirates, pp. 45–52.

Sheridan, P. K., G. Evans, and D. Reeve. 2014. "*Understanding Teaching Assistants' Assessment of Individual Teamwork Performance*." 121st ASEE Annual conference & exposition, Indianapolis, Indiana.

Shields, E. 2007. "*Fostering Creativity in the Capstone Engineering Design Experience*." 2007 ASEE Annual conference and exposition, Honolulu, Hawaii.

Shuman, L. J., M. Besterfield-Sacre, and J. McGourty. 2005. "The ABET "Professional Skills"—Can They be Taught? Can They be Assessed?" *Journal of Engineering Education* 94 (1): 41–55.

Solana-Gutiérrez, J., M.D. Bejarano-Carrión, and S. Merino-De-Miguel, "Triggering Student Creativity Through Statistical Brainstorming–an Experience in a Forestry Engineering Degree", 6th International conference on education and New learning technologies, Barcelona, Spain, 2014, pp. 6461-6470.

Strauss, L. C., and P. T. Terenzini. 2005. "*Assessing Student Performance on EC2000 Criterion 3. ak*." 2005 ASEE Annual conference and exposition, Portland, Oregon, pp. 1–12.

UNESCO. 2010. *Engineering: Issues, Challenges and Opportunities for Development*. Paris, France: UNESCO publishing.

Uruburu Colsa, Á, A. González Marcos, F. Alba Elias, and J. Ordieres Meré. 2011. "*Competence Monitoring in Project Teams by Using Web Based Portfolio Management Systems*." Proceedings of research in engineering education Symposium 2011, Madrid, Spain.

Valdes-Vasquez, R., and C. M. Clevenger. 2015. "Piloting Collaborative Learning Activities in a Sustainable Construction Class." *International Journal of Construction Education and Research* 11 (2): 79–96.

Vicent, L., J. Margalef, J. A. Montero, and J. Anguera. 2007. "*Work in Progress-Developing Transversal Skills in an Online Theoretical Engineering Subject*." 37th ASEE/IEEE Frontiers in education conference, Milwaukee, Wisconsin, pp. F3B-11-F13B-12.

Vila-Parrish, A., T. B. Baldwin, L. Battestilli, H. Queen, J. Y. Schmidt, and S. Carson. 2016. "*TH!NK: A Framework to Assess and Support Critical and Creative Thinking*." 123rd ASEE Annual conference & exposition, New Orleans, Louisiana.

Villiers, C., D. Bondehagen, and Y. Mehta. 2013. "*Assessment of a New Approach to Implement Life-Long Learning Into Civil and Environmental Engineering Courses*." 120th ASEE Annual conference & exposition, Atlanta, Louisiana.

Ward, M., L. Gruppen, and G. Regehr. 2002. "Measuring Self-Assessment: Current State of the art." *Advances in Health Sciences Education* 7 (1): 63–80.

Wasserman, J., M. Schmidt, and R. Jendrucko. 2003. "*The In-Class Use of Assessment and Rubrics by Student Groups to Improve Presentation Performance in Biomedical Engineering*." 2003 ASEE Annual conference & exposition.

Waychal, P. K. 2014. "*Developing Creativity Competency of Engineers*." 121st ASEE Annual conference & exposition, Indianapolis, Indiana.

Wertz, R. E. H., M. Fosmire, S. Purzer, A. I. Saragih, A. van Epps, M. R. Sapp Nelson, and B. G. Dillman. 2013. "*Work in Progress: Critical Thinking and Information Literacy: Assessing Student Performance*." 120th ASEE conference & exposition, Atlanta, Georgia.

Wertz, R. E. H., M. Ross, S. Purzer, M. Fosmire, and M. Cardella. 2011. "*Assessing Engineering Students' Information Literacy Skills: An Alpha Version of a Multiple-Choice Instrument*." 118th ASEE conference & exposition, Vancouver, Canada.

Whitmore, P. G., and J. P. Fry. 1974. *Soft Skills: Definition, Behavioral Model Analysis, Training Procedures*. Alexandria, VA: Ft. Belvoir Defense Technical Information Center MAR.: Human Resources Research Organization.

Wilkins, K. G., B. L. Bernstein, and J. M. Bekki. 2015. "Measuring Communication Skills: The STEM Interpersonal Communication Skills Assessment Battery." *Journal of Engineering Education* 104 (4): 433–453.

Wu, T., Y. Huang, and R. Shadiev. 2016. "*The Application of Multi-Dimensional Learning Portfolios for Exploring the Creativity Learning Behavior in Engineering Education*." 123rd ASEE Annual conference & exposition, New Orleans, Louisiana.

Zhang, A. 2012. "Peer Assessment of Soft Skills and Hard Skills." *Journal of Information Technology Education: Research* 11 (1): 155–168.

Zheng, W., J. Yin, H. Shi, and G. Skelton. 2016. "Prompted Self-Regulated Learning Assessment and Its Effect for Achieving ASCE Vision 2025." *Journal of Professional Issues in Engineering Education and Practice* 143 (2): 04016021.

Appendices

Appendix A. Description of communication criteria used in literature and the corresponding studies

Category	Criteria	Description	References
Oral Communication ($n = 31$)	Organization ($n = 15$)	Give a clear, logic, structured and organised oral presentation.	(Saunders et al. 2003; Wasserman, Schmidt, and Jendrucko 2003; Huyck, Ferguson, and Wasserman 2007; Amante et al. 2010; Keshavarz and Vaziri 2012; Ryan 2012; Jones and Abdallah 2013; Bousaba et al. 2014; Mourtos 2014; Colsa et al. 2015; Eichelman, Clark, and Bodnar 2015; García et al. 2015; Ortiz et al. 2015; de-Juan et al. 2016; Galván-Sánchez et al. 2017)
	Visual/written aids ($n = 11$)	Use media, graphics, writing texts that support the oral exposition	(Saunders et al. 2003; Huyck, Ferguson, and Wasserman 2007; Keshavarz and Vaziri 2012; Ryan 2012; Jones and Abdallah 2013; Bousaba et al. 2014; Mourtos 2014; Colsa et al. 2015; García et al. 2015; Ortiz et al. 2015; Galván-Sánchez et al. 2017)
	Speech delivery ($n = 9$)	Modulate the voice, maintain eye contact and use body language	(Saunders et al. 2003; Wasserman, Schmidt, and Jendrucko 2003; Amante et al. 2010; Keshavarz and Vaziri 2012; Ryan 2012; Bousaba et al. 2014; Colsa et al. 2015; Eichelman, Clark, and Bodnar 2015; Galván-Sánchez et al. 2017)
	Content development ($n = 7$)	Use orally appropriate, relevant and compelling content that shows mastery of the subject	(Saunders et al. 2003; Wasserman, Schmidt, and Jendrucko 2003; Ryan 2012; Jones and Abdallah 2013; Bousaba et al. 2014; Eichelman, Clark, and Bodnar 2015; García et al. 2015)
	Time management ($n = 7$)	Adhere to the presentation time limit	(Huyck, Ferguson, and Wasserman 2007; Amante et al. 2010; Keshavarz and Vaziri 2012; Ryan 2012; Bousaba et al. 2014; Mourtos 2014; Galván-Sánchez et al. 2017)
	Audience ($n = 4$)	Interact with and adapt the oral communication to different audiences	(Wasserman, Schmidt, and Jendrucko 2003; Bousaba et al. 2014; Dohaney et al. 2015; Ortiz et al. 2015)
	Questions and answers ($n = 4$)	Understand and answer questions	(Saunders et al. 2003; Amante et al. 2010; Keshavarz and Vaziri 2012; Ryan 2012)
	Team uniformity ($n = 2$)	Show homogeneous transition between the presenters (visual and oral presentation)	(Ryan 2012; Galván-Sánchez et al. 2017)
Written communication ($n = 24$)	Language ($n = 13$)	Use the appropriate vocabulary, syntax and grammar	(Saunders et al. 2003; Wasserman, Schmidt, and Jendrucko 2003; Flateby and Fehr 2008; Keshavarz and Vaziri 2012; Jones and Abdallah 2013; Mourtos 2014; Eichelman, Clark, and Bodnar 2015; Frank et al. 2015; García et al. 2015; Backer 2016; Barr 2016; de-Juan et al. 2016; Fries et al. 2017)
	Report organization ($n = 12$)	Develop a clear, logic, structured and organised report	(Saunders et al. 2003; Wasserman, Schmidt, and Jendrucko 2003; Flateby and Fehr 2008; Keshavarz and Vaziri 2012; Jones and Abdallah 2013; Mourtos 2014; Eichelman, Clark, and Bodnar 2015; Frank et al. 2015; Barr 2016; de-Juan et al. 2016; Ramírez-Echeverry, Olarte Dussán, and García-Carillo 2016; Fries et al. 2017)

(*Continued*)

Continued.

Category	Criteria	Description	References
	Content development ($n = 10$)	Write appropriate, relevant and compelling content that shows mastery of the subject	(Saunders et al. 2003; Wasserman, Schmidt, and Jendrucko 2003; Flateby and Fehr 2008; Keshavarz and Vaziri 2012; Jones and Abdallah 2013; Mourtos 2014; Eichelman, Clark, and Bodnar 2015; Frank et al. 2015; Barr 2016; Fries et al. 2017)
	Source citation ($n = 8$)	Cite sources accurately	(Saunders et al. 2003; Barnes, Dyrenfurth, and Newton 2012; Jones and Abdallah 2013; Eichelman, Clark, and Bodnar 2015; Frank et al. 2015; García et al. 2015; Barr 2016; Ramírez-Echeverry, Olarte Dussán, and García-Carillo 2016)
	Context and purpose ($n = 6$)	Perform a report according to the context and purpose of the situation	(Wasserman, Schmidt, and Jendrucko 2003; Flateby and Fehr 2008; Eichelman, Clark, and Bodnar 2015; Frank et al. 2015; Barr 2016; Ramírez-Echeverry, Olarte Dussán, and García-Carillo 2016)
	Audience ($n = 4$)	Adapt the written communication according to the audience	(Wasserman, Schmidt, and Jendrucko 2003; Flateby and Fehr 2008; Frank et al. 2015; Ramírez-Echeverry, Olarte Dussán, and García-Carillo 2016)
	Graphical/tables support ($n = 3$)	Use the appropriate graphics and tables to support the text	(Saunders et al. 2003; Jones and Abdallah 2013; Mourtos 2014)
Foster communication ($n = 7$)	Self-confidence ($n = 4$)	Act confidently in different situations	(Amante et al. 2010; Lopes et al. 2015; Ngaile, Wang, and Gau 2015; Wilkins, Bernstein, and Bekki 2015)
	Achieve/convey ideas ($n = 3$)	Achieve conversation goals, negotiate, persuade and convey ideas in different ways (e.g. graphically, verbally)	(Immekus et al. 2005; Strauss and Terenzini 2005; Hernandez-Linares et al. 2015)
	Self-exposure ($n = 2$)	Self-expose to conversations in different environments	(Lopes et al. 2015; Wilkins, Bernstein, and Bekki 2015)
Foster interaction ($n = 4$)	Group interaction ($n = 3$)	Interact with all group members and encourage communication between members	(Ater Kranov et al. 2008; Uruburu Colsa et al. 2011; Schmeckpeper et al. 2014)
	Group self-regulation ($n = 3$)	Work together to find consensus considering perspectives	(Ater Kranov et al. 2008; Kaul and Adams 2014; Schmeckpeper et al. 2014)
	Client interaction ($n = 1$)	Interact and communicate with clients or instructors	(Uruburu Colsa et al. 2011)
Communication management ($n = 1$)	Conflict resolution ($n = 1$)	Use communication in order to avoid and solve conflicts	(Immekus et al. 2005)
Reading ($n = 1$)	Reading ($n = 1$)	Read and understand the reading	(Backer 2016)
Listening ($n = 2$)	Listening ($n = 2$)	Listen, understand and react on verbal messages	(Mourtos 2014; Backer 2016)

Appendix B. Description of lifelong learning criteria used in literature

Category	Criteria	Description	References
Reflection (*n* = 21)	Self-reflection (*n* = 17)	Reflect on current knowledge and to differentiate from what is known or not known	(Mourtos 2003; Ater Kranov et al. 2008; Kirby et al. 2010; Ater Kranov et al. 2011; Lord et al. 2011; Martínez-Mediano and Lord 2012; Ater Kranov et al. 2013; Chen, Lord, and McGaughey 2013; Douglas et al. 2014; Kaul and Adams 2014; Mourtos 2014; Schmeckpeper et al. 2014; Frank et al. 2015; Ortiz et al. 2015; Schoepp, Danaher, and Ater Kranov 2016; Zheng et al. 2016; Ahmed 2017)
	Self-monitoring (*n* = 6)	Monitor learning process and self-assess periodically the performance	(Ater Kranov et al. 2008; Ater Kranov et al. 2011; Ater Kranov et al. 2013; Schmeckpeper et al. 2014; Schoepp, Danaher, and Ater Kranov 2016; Zheng et al. 2016)
	Goal setting (*n* = 5)	Set goals for tasks that might be important for the development	(Kirby et al. 2010; Lord et al. 2011; Chen, Lord, and McGaughey 2013; Ortiz et al. 2015; Zheng et al. 2016)
	Responsible for own learning (*n* = 4)	Monitor and evaluate individual and others learning	(Litzinger et al. 2001; Briedis 2002; Martínez-Mediano and Lord 2012; Ahmed 2017)
	Critical thinking (*n* = 4)	Think carefully about a subject considering various perspectives to arrive at an appropriate solution or conclusion	(Litzinger et al. 2001; Lord et al. 2011; Villiers, Bondehagen, and Mehta 2013; Wertz et al. 2013)
	Learn from mistakes (*n* = 1)	Learn from mistakes and to practice continuous improvement	(Briedis 2002)
Acquiring (*n* = 20)	Locating and scrutinising information (*n* = 16)	Locate, examine closely and interpret information	(Ater Kranov et al. 2008; Kirby et al. 2010; Ater Kranov et al. 2011; Lord et al. 2011; Wertz et al. 2011; Ater Kranov et al. 2013; Chen, Lord, and McGaughey 2013; Villiers, Bondehagen, and Mehta 2013; Wertz et al. 2013; Douglas et al. 2014; Scharf 2014; Schmeckpeper et al. 2014; Ortiz et al. 2015; Schoepp, Danaher, and Ater Kranov 2016; Zheng et al. 2016; Ahmed 2017)
	Source citation (*n* = 4)	Identify and cite correctly quality sources	(Wertz et al. 2011; Wertz et al. 2013; Douglas et al. 2014; Scharf 2014)
	Reading engineering information (*n* = 2)	Take the initiative to read engineering articles or books outside of classes	(Mourtos 2003; Mourtos 2014)
	Apply technology (*n* = 2)	Acquire and apply new technologies	(Strauss and Terenzini 2005; Douglas et al. 2014)
Learning (*n* = 14)	Willingness/ motivation/ curiosity to learn (*n* = 11)	Learn new things autonomously	(Litzinger et al. 2001; Briedis 2002; Mourtos 2003; Strauss and Terenzini 2005; Coşkun and Demirel 2010; Lord et al. 2011; Martínez-Mediano and Lord 2012; Kaul and Adams 2014; Mourtos 2014; Frank et al. 2015; Ortiz et al. 2015)
	Rehearse and elaborate information (*n* = 3)	Memorize, summarise and paraphrase information	(Lord et al. 2011; Scharf 2014; Frank et al. 2015)
	Adaptation of learning strategies (*n* = 3)	Deal with uncertain and unexpected conditions and to solve problems when they arise	(Kirby et al. 2010; Chen, Lord, and McGaughey 2013; Ortiz et al. 2015)
	Creativity (*n* = 1)	Create new ideas	(Litzinger et al. 2001)
Initiating (*n* = 11)	Create a learning plan (*n* = 10)	Prepare own personal learning development plan to initiate tasks or participate in activities to improve future performance	(Litzinger et al. 2001; Mourtos 2003; Kirby et al. 2010; Martínez-Mediano and Lord 2012; Chen, Lord, and McGaughey 2013; Villiers, Bondehagen, and Mehta 2013; Douglas et al. 2014; Mourtos 2014; Frank et al. 2015; Ahmed 2017)
	Help-seeking (*n* = 2)	Seek help when needed	(Lord et al. 2011; Frank et al. 2015)
Participating (*n* = 5)	Activity participation (*n* = 4)	Participate in professional activities and learning opportunities	(Litzinger et al. 2001; Briedis 2002; Mourtos 2003; Mourtos 2014)
	Collaborative work (*n* = 1)	Collaborate in teams to solve complex problems	(Martínez-Mediano and Lord 2012)

Appendix C. Description of teamwork criteria used in literature.

Categories	Criteria	Description	References
Team interaction ($n = 26$)	Interacting with others ($n = 18$)	Interact and work with others to accomplish team goals and to solve problems	(Immekus et al. 2005; Strauss and Terenzini 2005; Besterfield-Sacre et al. 2007; Huyck, Ferguson, and Wasserman 2007; Pazos, Micari, and Light 2010; Gerlick et al. 2011; Uruburu Colsa et al. 2011; Fini and Mellat-Parast 2012; Zhang 2012; Enszer and Castellanos 2013; Sheridan, Evans, and Reeve 2014; Gerhart and Carpenter 2015; Ortiz et al. 2015; Valdes-Vasquez and Clevenger 2015; Brake and Curry 2016; Bringardner et al. 2016; Pazos et al. 2016; Blanco, López-Forniés, and Zarazaga-Soria 2017)
	Contribution of ideas/solutions/ work ($n = 9$)	Contribute with ideas, solutions and effort in a group work	(Besterfield-Sacre et al. 2007; Amante et al. 2010; Gerlick et al. 2011; Fini and Mellat-Parast 2012; Zhang 2012; Enszer and Castellanos 2013; Fagerholm and Vihavainen 2013; Sheridan, Evans, and Reeve 2014; Valdes-Vasquez and Clevenger 2015)
	Problem-solving and decision making ($n = 8$)	Make decisions and solve problems respecting the team consensus	(Saunders et al. 2003; Besterfield-Sacre et al. 2007; Huyck, Ferguson, and Wasserman 2007; Vicent et al. 2007; Pazos, Micari, and Light 2010; Fini and Mellat-Parast 2012; Kaul and Adams 2014; Mourtos 2014)
	Multidisciplinary work ($n = 5$)	Work in teams of people of a variety of skills and backgrounds	(Strauss and Terenzini 2005; Uruburu Colsa et al. 2011; Kaul and Adams 2014; Mourtos 2014; Blanco, López-Forniés, and Zarazaga-Soria 2017)
	Proactivity to collaborate/ Participate ($n = 4$)	Actively participate and collaborate with team members	(Uruburu Colsa et al. 2011; Zhang 2012; García et al. 2015; Hernandez-Linares et al. 2015)
	Prepared for team meetings ($n = 4$)	Attend and be prepared for team meetings	(Saunders et al. 2003; Huyck, Ferguson, and Wasserman 2007; Zhang 2012; Sheridan, Evans, and Reeve 2014)
	Engineering discipline work ($n = 1$)	Work in teams of people of different engineering disciplines	(Strauss and Terenzini 2005)
Work management ($n = 20$)	Manage team responsibility ($n = 15$)	Define team roles, plan work, assign tasks, distribute work equally among team members and set direction	(Saunders et al. 2003; Immekus et al. 2005; Besterfield-Sacre et al. 2007; Huyck, Ferguson, and Wasserman 2007; Vicent et al. 2007; Pazos, Micari, and Light 2010; Gerlick et al. 2011; Zhang 2012; Fagerholm and Vihavainen 2013; Jones and Abdallah 2013; Sheridan, Evans, and Reeve 2014; Dunai et al. 2015; Brake and Curry 2016; Bringardner et al. 2016; Pazos et al. 2016)
	Monitored and controlled activities ($n = 6$)	Monitor and control team activities to fulfil goals and deadlines	(Saunders et al. 2003; Huyck, Ferguson, and Wasserman 2007; Mourtos 2014; Garcia et al. 2015; Valdes-Vasquez and Clevenger 2015; Pazos et al. 2016)
	Responsible for team outcomes ($n = 6$)	Display dedication, determination and responsibility on team outcomes and performance	(Huyck, Ferguson, and Wasserman 2007; Dolan et al. 2011; Zhang 2012; Fagerholm and Vihavainen 2013; Sheridan, Evans, and Reeve 2014; Bringardner et al. 2016)
	Team leadership ($n = 6$)	Demonstrate leadership by taking responsibility for tasks and by motivating and disciplining others	(Dolan et al. 2011; Zhang 2012; Gerhart and Carpenter 2015; Brake and Curry 2016; Bringardner et al. 2016; Wu, Huang, and Shadiev 2016)
Fostering team climate ($n = 21$)	Team relationship ($n = 15$)	Engage members with respect and listen and respect others' ideas	(Saunders et al. 2003; Besterfield-Sacre et al. 2007; Dolan et al. 2011; Gerlick et al. 2011; Fini and Mellat-Parast 2012; Zhang 2012; Fagerholm and Vihavainen 2013; Jones and Abdallah 2013; Sheridan, Evans, and Reeve 2014; Dunai et al. 2015; Garcia et al. 2015; Valdes-Vasquez and Clevenger 2015; Bringardner et al. 2016; Heinis, Goller, and Meboldt 2016; Blanco, López-Forniés, and Zarazaga-Soria 2017)

(*Continued*)

Continued.

Categories	Criteria	Description	References
	Encourage the group ($n = 7$)	Encourage the group to achieve goals and to contribute to group work	(Huyck, Ferguson, and Wasserman 2007; Amante et al. 2010; Gerlick et al. 2011; Zhang 2012; Enszer and Castellanos 2013; Sheridan, Evans, and Reeve 2014; Valdes-Vasquez and Clevenger 2015)
	Managing conflict ($n = 5$)	Avoid conflict and to solve conflicts	(Gerlick et al. 2011; Mourtos 2014; Brake and Curry 2016; Bringardner et al. 2016; Pazos et al. 2016)
Foster interaction ($n = 13$)	Communication with group members/others ($n = 9$)	Communicate actively and constructively between group members and others outside the discipline	(Besterfield-Sacre et al. 2007; Ramirez, Jimenez, and Hernandez 2007; Amante et al. 2010; Pazos, Micari, and Light 2010; Gerlick et al. 2011; Jones and Abdallah 2013; Dunai et al. 2015; Valdes-Vasquez and Clevenger 2015; Pazos et al. 2016)
	Stimulate industry's involvement ($n = 3$)	Stimulate the involvement and communication with industry partners	(Gerlick et al. 2011; Mourtos 2014; Garcia et al. 2015)
	Accept/ask feedback ($n = 2$)	Accept or ask feedback to improve personally or the team performance	(Kaul and Adams 2014; Sheridan, Evans, and Reeve 2014)
	Provide opinions and constructive feedback ($n = 1$)	Provide opinions and constructive feedback to others	(Sheridan, Evans, and Reeve 2014)
Work delivery ($n = 9$)	Work content ($n = 6$)	Develop a team work with quality content	(Gerlick et al. 2011; Zhang 2012; Enszer and Castellanos 2013; Sheridan, Evans, and Reeve 2014; Valdes-Vasquez and Clevenger 2015; Wu, Huang, and Shadiev 2016)
	Proficiency in discipline/topic ($n = 3$)	Show proficiency in the discipline or topic developed	(Zhang 2012; Enszer and Castellanos 2013; Mourtos 2014)
	Homogenous report/ presentation ($n = 2$)	Elaborate a homogenous presentation and report when the individual parts are combined	(Huyck, Ferguson, and Wasserman 2007; Ortiz et al. 2015)
	Presentation and explanation of work ($n = 2$)	Present, explain and defend the work developed	(Huyck, Ferguson, and Wasserman 2007; Valdes-Vasquez and Clevenger 2015)

Appendix D. Description of innovation criteria used in literature

Categories	Criteria	Description	References
Idea generation (*n* = 17)	Flexibility (*n* = 15)	Apply and integrate engineering knowledge, and to use different technologies to generate ideas and solutions	(Ramirez, Jimenez, and Hernandez 2007; Shields 2007; Charyton, Jagacinski, and Merrill 2008; Hernández and Ramírez 2008; Robbins and Kegley 2010; Charyton et al. 2011; Amelink et al. 2013; Solana-Gutiérrez, Bejarano-Carrión, and Merino-De-Miguel 2014; Hernandez-Linares et al. 2015; Clemente, Vieira, and Tschimmel 2016; Lopez-Malo et al. 2016; Vila-Parrish et al. 2016; Wu, Huang, and Shadiev 2016; Keh, Ismail, and Yusof 2017; Saorín et al. 2017)
	Originality (*n* = 13)	Produce unique or unusual ideas in a given context	(Shields 2007; Charyton, Jagacinski, and Merrill 2008; Robbins and Kegley 2010; Charyton et al. 2011; Eppes, Milanovic, and Sweitzer 2012; Fila and Purzer 2012; Solana-Gutiérrez, Bejarano-Carrión, and Merino-De-Miguel 2014; Clemente, Vieira, and Tschimmel 2016; Lopez-Malo et al. 2016; Vila-Parrish et al. 2016; Wu, Huang, and Shadiev 2016; Keh, Ismail, and Yusof 2017; Saorín et al. 2017)
	Fluency (*n* = 7)	Produce a great number of ideas and solutions	(Shields 2007; Charyton, Jagacinski, and Merrill 2008; Robbins and Kegley 2010; Charyton et al. 2011; Solana-Gutiérrez, Bejarano-Carrión, and Merino-De-Miguel 2014; Wu, Huang, and Shadiev 2016; Keh, Ismail, and Yusof 2017)
	Elaboration (*n* = 7)	Pull information together, add details, and explain and polish information	(Amelink et al. 2013; Solana-Gutiérrez, Bejarano-Carrión, and Merino-De-Miguel 2014; Clemente, Vieira, and Tschimmel 2016; Lopez-Malo et al. 2016; Vila-Parrish et al. 2016; Wu, Huang, and Shadiev 2016; Saorín et al. 2017)
Product generation (*n* = 10)	Connectivity (*n* = 4)	Connect and integrate different ideas and solutions into new creative forms	(Eppes, Milanovic, and Sweitzer 2012; Garcia Garcia et al. 2014; Clemente, Vieira, and Tschimmel 2016; Saorín et al. 2017)
	Implementing solutions (*n* = 3)	Work with team members to design and implement innovative solutions	(Ramirez, Jimenez, and Hernandez 2007; Hernández and Ramírez 2008; Amelink et al. 2013)
	Usefulness (*n* = 3)	Develop a product or system that fulfils the given design problem	(Charyton et al. 2011; Fila and Purzer 2012; Lopez-Malo et al. 2016)
	Viability (*n* = 1)	Develop a product or system that can be maintained and sustained	(Fila and Purzer 2012)
	Feasibility (*n* = 1)	Develop a product or system that can be implemented	(Fila and Purzer 2012)
	Desirability (*n* = 1)	Develop a product that is accepted by consumers and users	(Fila and Purzer 2012)
Thinking (*n* = 9)	Scaling information (*n* = 4)	Identify, organise and synthesise information, elements and ideas	(Eppes, Milanovic, and Sweitzer 2012; Amelink et al. 2013; Hernandez-Linares et al. 2015; Lopez-Malo et al. 2016)
	Critical thinking (*n* = 2)	Think carefully, explore and evaluate ideas considering various perspectives to arrive an appropriate solution or conclusion	(Amelink et al. 2013; Vila-Parrish et al. 2016)
	Problem-solving (*n* = 2)	Develop a logical and consistent plan to solve problems and to implement it effectively	(Eppes, Milanovic, and Sweitzer 2012; Lopez-Malo et al. 2016)
	Abstract thinking (*n* = 1)	Think in terms of concepts and general principals	(Garcia Garcia et al. 2014)
Communication (*n* = 3)	Graph ability (*n* = 2)	Express graphically	(Clemente, Vieira, and Tschimmel 2016; Saorín et al. 2017)
	Communicate innovation (*n* = 1)	Communicate innovative designs and solutions	(Amelink et al. 2013)

(*Continued*)

Continued.

Categories	Criteria	Description	References
Risk ($n = 3$)	Risk-taking ($n = 3$)	Take personal risk or risk of product failure	(Eppes, Milanovic, and Sweitzer 2012; Lopez-Malo et al. 2016; Vila-Parrish et al. 2016)
Contradiction ($n = 3$)	Embracing contradiction ($n = 3$)	Recognize the value of divergent or contradictory perspectives and to integrate them	(Eppes, Milanovic, and Sweitzer 2012; Lopez-Malo et al. 2016; Vila-Parrish et al. 2016)
Active learning ($n = 3$)	Acquiring competencies ($n = 2$)	Attain strategies and skills within a particular domain	(Eppes, Milanovic, and Sweitzer 2012; Lopez-Malo et al. 2016)
	Self-initiated exploration ($n = 1$)	Direct monitor own learning and knowledge	(Amelink et al. 2013)
Diffusion ($n = 2$)	Diffusion ($n = 2$)	Explore the resources that lead to participation in innovative activities or development of innovation	(Amelink et al. 2013; Hernandez-Linares et al. 2015)
Collaboration ($n = 1$)	Collaboration ($n = 1$)	Work with others to generate ideas	(Amelink et al. 2013)

Appendix E. Reliability and validity evidence of the methods. (IRR indicates interrater reliability; ICC - intra-class correlation coefficient; r - Pearson correlation coefficient; EFA - exploratory factor analysis; CFA - confirmatory factor analysis; ρ - Spearman correlation; PA - percentage agreement; KMO - Kaiser Meyer Olkin)

Instrument, Dimensions, Scale	Competency	Reliability	Validity	Reference
Rubric 14 dimensions 3-point Likert scale	Communication (Oral)	Internal consistency (2 raters): Cronbach's α=0.86 and 0.74 IRR (2 raters): r = 0.68 (p < 0.001) and Cohen's kappa statistic (strict criterion = 0.33 and lax criterion = 0.40)	Content validity: literature and expert review Tested as pilot study	(Galván-Sánchez et al. 2017)
Social Skill Inventory (SSI-Del-Prette) 38 dimensions	Communication	Internal consistency: Cronbach's α=0.75 Test-retest reliability: r = 0.90 (p < 0.001)	Pre and post-assessment	(Lopes et al. 2015)
Science, Technology, Engineering and Math Interpersonal Communication Skills Assessment Battery (STEM ICSAB): 3 instruments of 20, 12 and 5 dimensions	Communication	Internal consistency: EFA: Cronbach's α=0.96, 0.92 and 0.85 CFA: Cronbach's α=0.93, 0.94 and 0.66	Content validity: expert review Construct validity: EFA: 1 factor solution VarExp: 53%, 57%, 46% FL ≥ 0.60, 0.60, 0.60 CFA: 1 factor solution FL ≥ 0.55, 0.54, 0.38	(Wilkins, Bernstein, and Bekki 2015)
Cognitive Level and Quality of Writing Assessment, CLAQWA	Communication (Written)	IRR: Correlations above 0.80	Content validity: Expert review/validation 6-year study	(Flateby and Fehr 2008)
Written VALUE rubric 4 Likert scale Elevator Pitch 5 Likert scale	Communication (Written)	IRR (2 raters) ICC(VALUE rubric) = 0.73 (individual dimensions - 0.56–0.73) ICC (Elevator Pitch) = 0.59 (individual dimensions - from 0.64–0.75)	Control and experimental groups Construct validity: correlation between students grades and their perception questionnaire of communication improvement (for experimental group r=−0.11, ρ=−0.04, p = 0.86, and for control group r = 0.04, ρ=0.04, p = 0.88)	(Eichelman, Clark, and Bodnar 2015)
Collegiate Learning Assessment, CLA+	Communication (Written)	Internal consistency: Stratified α=0.85 and 0.87	Face validity: students self-report on what skills they perceived the test measured Construct validity: correlation with 2 measures (CLA Performance Task r = 0.73-0.83 and CLA Critique an Argument r = 0.73-0.94)	(Frank et al. 2015)
Written VALUE rubric 5 dimensions 4 Likert scale	Communication (Written)	IRR (2 raters): Kappa statistics >90%	Face validity	(Frank et al. 2015)
Post-questionnaire 36 dimensions (CM 4 items, TW 3 items, and LLL 3 items) 5 Likert scale	Communication Lifelong learning Teamwork	Internal consistency: Cronbach's α(CM) = 0.86, α(LLL) = 0.78, and α(TW) = 0.86	Tested as pilot test Content validity: Review of literature and available instruments, and experts and students review of item and content Construct validity: EFA 9 factors FL (CM) ≥ 0.60, FL(LLL) ≥ 0.73, FL(TW) ≥ 0.65 ExpVar: 72.2%	(Strauss and Terenzini 2005)
Questionnaire 23 dimensions 5 Likert scale	Communication Teamwork	Internal consistency: Cronbach's α(TW) = 0.905 and 0.94, α(CM) = 0.857 and 0.651	Pre and post-questionnaire 5 factors	(Fini and Mellat-Parast 2012)
				(Immekus et al. 2005)

(*Continued*)

Continued.

Instrument, Dimensions, Scale	Competency	Reliability	Validity	Reference
Self-report CM 26 items, and TW 15 items) 5 Likert scale	Communication Teamwork	Internal consistency: Cronbach's α(TW) = 0.94, α(CM) = 0.95	Pilot study is described Content validity: review of literature and items Construct validity: CFA FL(TW) > 0.22 and FL(CM) > 0.20	
Questionnaire of 45 dimensions (CM - 5 items, IC - 3 items for creativity, and TW - 1 item) 5 Likert scale	Communication Innovation	Internal consistency: Cronbach's α=0.968 α (CM) = 0.823, α(I) = 0.663	Validated in previous work Construct validity: EFA (previous work) and CFA 9 factors ExpVar: 74.109%	(Hernandez-Linares et al. 2015)
Modified Strategies for Learning Questionnaire, MSLQ 5 dimensions 5 Likert scale	Innovation	Internal consistency (calculated in previous study): Moderate Cronbach's α<0.70	Predicted validity: scales related to academic performance	(Lord et al. 2011; Amelink et al. 2013)
Abreaction Test for Evaluating Creativity, TAEC	Creativity	Equivalent forms: Wilcoxon ($p < 0.01$ for experimental group)	Pre and post-intervention Control and experimental groups	(Clemente, Vieira, and Tschimmel 2016)
Creative Thinking VALUE rubric 6 dimensions 4 Likert scale	Creative Thinking	IRR (2 raters): 0.85 and 0.79		(Lopez-Malo et al. 2016)
Abreaction Test of Creativity 12 dimensions 5 Likert scale	Innovation	Internal consistency: Cronbach's α=0.71		(Saorín et al. 2017)
Critical Thinking Assessment Test, CAT	Critical and creative thinking	IRR: average score = 0.82	Face and criterion validity (correlation between 5 measurements) Pre and post-test	(Vila-Parrish et al. 2016)
Scoring rubric 5 dimensions 5 Likert scale	Innovation	IRR(2 raters): ρ ($p < 0.05$ for all 4 dimensions)	Content validity (literature review)	(Fila and Purzer 2012)
Index of Learning Styles, ILS 4 dimensions	Creativity	Internal consistency: Cronbach's α of moderate effect (calculated in previous studies)	Pre and post-assessment Convergent validity	(Waychal 2014)
Torrance Tests of Creativity Thinking 4 dimensions	Innovation	Reliability assessed in previous studies	Predicted validity	(Shields 2007; Robbins and Kegley 2010; Wu, Huang, and Shadiev 2016)
Creative engineering Design Assessment, CEDA 4 dimensions	Creativity	IRR (2 raters): r = 0.98 Consistency of pre and post-test reliability: r = 0.563 CEDA was moderately correlated with the PCT (r = 0.39, $p < 0.01$) and slightly correlated with the PVST-R (r = 0.19, $p < 0.05$)	Convergent validity (correlation between 3 instruments) Discriminant validity	(Charyton, Jagacinski, and Merrill 2008; Charyton et al. 2011)
Lifelong Learning Scale, LLS 14 dimensions	Lifelong Learning	Internal consistency: Cronbach's α(LLL) = 0.71 57% factor loadings > 0.4		(Kirby et al. 2010; Chen, Lord, and McGaughey 2013)
Critical Engineering Literacy Test, CELT, 28 dimensions	Lifelong Learning	Internal consistency: Cronbach's α(LLL) = 0.67; r = 0.47 ($p < 0.01$)	Construct validity (Correlation between CELT and Critical Assessment Test)	(Wertz et al. 2013)
Engineering Professional Skills Assessment, EPSA	Lifelong Learning	IRR: inter-rater agreement	Literature review and expert review of content, construct and criterion validity Pilot study Control group	(Ater Kranov et al. 2008; Ater Kranov et al. 2011; Ater Kranov et al. 2013; Schmeckpeper et al. 2014)
Scharf Diagnostic Essay Prompt 5 dimensions	Lifelong Learning	IRR: high inter-item correlation matrix Internal consistency: Cronbach's α=0.959	Predicted validity (the overall grades correlate with pre and post-essay scores)	(Scharf 2014)
Self-Assessment of Problem Solving Strategies, SAPSS 20 dimensions	Lifelong Learning	Internal consistency: Cronbach's α=0.92 Factor loadings ≥ 0.34	Content validity: literature review Construct validity (EFA)	(Douglas et al. 2014)

(*Continued*)

Continued.

Instrument, Dimensions, Scale	Competency	Reliability	Validity	Reference
Lifelong Tendency Scale, LLTS 27 dimensions	Lifelong Learning	Internal consistency: Cronbach's α=0.89 KMO = 0.89 Factor loadings $\geq$ 0.32	Content validity (literature review, then students and lecturers review, and test correlation with Index Curiosity) Construct validity: KMO coefficient, Barlett test and EFA	(Coşkun and Demirel 2010)
Instrument 10 dimensions	Teamwork	IRR: inter-class correlation coefficient = 0.93 Internal consistency: Cronbach's $\alpha \geq 0.76$ for 80% of the factors and $\alpha \geq 0.61$ for 20% of items	Content validity (literature review) Construct validity (EFA and CFA) Pilot study	(Pazos, Micari, and Light 2010)
Questionnaire 18 dimensions	Teamwork	Internal consistency: Cronbach's α(TW) $\geq$ 0.92 for all the factors $\alpha \geq 0.50$ for all the factors	Content validity (Literature review) CFA (all factor loadings > 0.5)	(Zhang 2012)
Team-Effective Inventory 27 dimensions	Teamwork	IRR (inter-class correlation)	Content validity (literature and items review, and experts review)	(Sheridan, Evans, and Reeve 2014)
Rubric 7 dimensions	Teamwork	IRR (4 raters): $r \geq 0.4$ for 4 dimensions and PA	Content validity (Experts and students review)	(Gerlick et al. 2011)
Questionnaire 8 dimensions	Teamwork	Internal consistency: Cronbach's $\alpha = 0.78$	Pre and post-assessment Group comparison	(Bringardner et al. 2016)
Work sampling observation	Teamwork	IRR	Valid	(Besterfield-Sacre et al. 2007)

*According to Field (2009):

The values of the factor loadings should be > 0.40 or Kaiser-Meyer-Olkin > 0.5, and the eigenvalues > 1.0.

IRR is assessed by intra-class correlation coefficient (ICC > 0 show similarities) and Pearson or Spearman rho correlation (absolute values closer to 1 indicate higher correlation).

Cronbach-α values around 0.80 are good for cognitive tests, around 0.70 for ability tests and in social sciences values below 0.7 can even be expected because of the diversity of the construct.