FACULTY OF SCIENCE

An in-depth study of professional identity of engineering students

Sybren DE BOEVER

Supervisor: Prof. Greet Langie
Faculty of Engineering Technology

Co-supervisor: Prof. An Carbonez
Leuven Statistics Research Centre (LStat)

Mentor: Sofie Craps
Faculty of Engineering Technology

Thesis presented in fulfillment of the requirements for the degree of Master of Science in Statistics and Data Science

Academic year 2020-2021

Acknowledgements

Exploration. Awareness. Confidence. These are three concepts that form the topic of this thesis, and which simultaneously closely represent the journey of this research. At each of these three stages, I had the privilege to collaborate with three power women - professor Langie, professor Carbonez, and doctor-to-be Sofie - who have guided and supported me through the ups and downs of this thesis. I am grateful for the opportunity to expand my horizon in this field of psychology and education which I had not explored before but eventually triggered my interest making this topic exciting to work on. I also explicitly want to thank professor Langie, professor Carbonez and Sofie for giving me the freedom to choose the direction I wanted to go in during this thesis, and for their accessibility and willingness to devote their time to this project. Additionally, I am very grateful that you made me aware of the SEFI conference and for allowing me to write a conference paper, which I consider to be the biggest opportunity and achievement of my student career. Finally, the positive critiques and support during this entire year helped me to develop my drive and confidence in this project. For all of this, I want to thank you! It was a joy to work with you.

Like in professional identity development, the support of my friends and family was of significant value during this journey. Thank you to my family for always taking an interest, for listening to my repetitive stories, and for always seeing the potential in me, even more than I can give myself credits for. Thank you to my friends for all the joyful moments that recharged my batteries. A special thank you to Marlies, Raf, Lawson and Maria for being my sideline cheerleaders during the entire master's programme and for your unconditional support and the ability to mentally transport me to another world. The evidence does support that the sushi dates, movie nights, walks and talks helped to make this work possible, which are experiments I gladly repeat in the future!

Index

Abstract

Engineering students often experience career orientation difficulties as a result of the numerous career options offered in the broad engineering field and/or a lack of professional and self-knowledge. This can contribute to a hampered transition from academia to the professional world and difficulties in career decision-making. Promoting a professional identity can improve the career orientation which can be achieved by career guidance that focusses on stimulating key constructs underlying the professional identity development. Because of its vast implications, professional identity has been investigated greatly, however, research focusing on engineering students is scarce. Additionally, researchers mainly focused on a restricted number of identity constructs, impairing a broad view of the professional identity in the same cohort. This study aimed to address these limitations by presenting an in-depth investigation of the professional identity of 624 engineering students of the Faculty of Engineering Technology at KU Leuven. This is achieved by examining five constructs simultaneously, including *career exploration*, *professional roles awareness*, *competence awareness*, *role-fit confidence*, and *competence confidence*. Using structural equation modeling techniques and survey data, we validated the survey design in measuring the constructs, we provided insights in the complex interplay between the identity constructs, and we determined several personal factors that altered the professional identity of engineering students. The results indicated the presence of a substantial positive correlation structure among the constructs, with possible stimulatory effects originating at *career exploration* that move downstream to *awareness* and *confidence*. Construct differences were observed for five of the six personal variables, i.e. phase of study, engineering persistence, professional role interest, migration status, and parental occupation, whereas no gender effects were present. These results contribute to a more general understanding of the professional identity of engineering students and, hence, provide fundamental support for the career guidance of these students during their education.

Keywords

Professional identity, engineering, structural equation modeling, career exploration, self-awareness, professional roles awareness, role-fit confidence, career guidance

Abbreviation list

CFA	Confirmatory factor analysis
CFI	Comparative fit index
DWLS	Diagonally weighted least squares
EFA	Exploratory factor analysis
FDR	False discovery rate
KMO	Kaiser-Meyer-Olkin
MIMIC	Multiple indicator multiple cause
MSA	Measure of sampling adequacy
RMSEA	Root mean squared error of approximation
SEM	Structural equation modeling
SRMR	Standardized root mean squared residuals
STEM	Science, Technology, Engineering and Mathematics
TLI	Tucker-Lewis index
ULS	Unweighted least squares

1. Introduction

1.1 Problem definition

Engineering students are presented many choices on a frequent basis that have mild to severe impact on their future (Matusovich *et al*., 2009). Of these, deciding on a specific future career at the beginning of their professional orientation is considered one of the most difficult and impactful decisions they have to face (Gati & Levin, 2014; Özek & Ferraris, 2018). Earlier in history, the career-decision making process was a less drastic event that was predominantly guided by gender stereotypes and familial inheritance to find a long-lasting job (Inkson & Elkin, 2008). Today, this ideology has faded because of the rapidly evolving and dynamic society that presents more possibilities and freedom (Savickas *et al*., 2009), which has led to several career orientation difficulties among engineering students, and students in general. A first challenge throughout engineering education is that the overload of available career information and the broad gamma of possible engineering careers can overwhelm students to an extent that it hampers their career orientation (Hofland *et al*., 2015; Gati & Levin, 2014). Secondly, students often perceive a low correspondence between their education and the engineering field (Trevelyan, 2019), or they fail to construct a clear view on what engineering entails (Bennett & Male, 2017; Matusovich *et al*., 2009; Saunders-Smits *et al*., 2021). Additionally, engineers have frequently shown to not persist in the engineering field, either during their education or after graduation (van Hattum-Janssen & Endedijk, 2020; Lichtenstein *et al*., 2009). These difficulties encountered by engineering students contribute to a difficult career orientation and hampered transition to the professional world.

These career orientation difficulties have been linked to a difficult professional identity development, which is a process where students become aware of their desires, needs and goals, but also their qualities, ambitions and how these coincide with the many professional possibilities in the job world (Kracke, 1997; Skorikov, 2011; Super, 1980). Because of its fundamental impact, the topic of professional identity has gained a large focus over the years with many studies contributing to the understanding of its development and implications (Skorikov, 2011), which provided essential insights for the design of career guidance programmes to assist students in their career orientation. Currently, the dual complexity of professional identity has been widely acknowledged following the identification of a variety of personal factors and multiple latent identity constructs that underly professional identity (Patrick & Borrego, 2016; Morelock, 2017). However, researchers have mainly focused on non-engineering students, thereby presenting the desire to improve the understanding of the professional identity of engineering students, which could provide important insights that could assist the development of career guidance for these students.

It is therefore the purpose of this thesis to help gain additional knowledge on several aspects of the professional identity of KU Leuven engineering students at the Faculty of Engineering

Technology. These insights can help shape the role of education in providing better support for these students in their transition from academia to the job world. Thereby, the career orientation difficulties associated with the engineering field could be tackled to improve for instance engineering persistence or career-fit. The following introductory sections will present some background on what professional identity entails, its importance, how it can be promoted, what it is influenced by, and gaps in the current identity literature are identified.

1.2 The concept of professional identity: what is it?

Professional identity development, also referred to as career, work, occupational or vocational identity development, is a process where students obtain information regarding themselves and the professional world by means of self-exploration, career information seeking and gaining real life experiences (Super, 1980). In this way, students create a career purpose along with a fundamental skillset that prepare them to comply and adopt better to the obligations of the professional world (Skorikov, 2011). This development reaches its significance when facing the transition to the work world (Özek & Ferraris, 2018), but it is generally seen as a complex lifelong process that starts already early on in life and maintains developing throughout education and adulthood (Schmitt-Rodemund, 1999; Super, 1980; Vondracek, 1992).

Researchers collectively agree about the complex nature of professional identity as several aspects have been identified that contribute to its development, including multiple underlying latent identity constructs and personal factors. Several of these key constructs are (1) '*career exploration*', representing the behavior of information seeking regarding the work world and oneself (Gagnon *et al*., 2019), which is considered fundamental for developing a professional identity (Super, 1980), (2) '*self-awareness*', representing a personal sense regarding one's aspirations, abilities and limitations (Watts, 1997), which allows students to establish professional goals and a purpose (Blustein, 1989; Kosby & Mariano, 2011), (3) '*career awareness*', presenting a professional sense about different job possibilities and values (Currie, 1975; Wise *et al*., 1976), and (4) '*career confidence*', indicating the confidence in one's skills for a career, i.e. *competence confidence*, and the confidence that a pursued career is suitable for oneself, i.e. *career-fit confidence* (Cech *et al*., 2011). These constructs, among others, constitute the complex multifaceted view of professional identity and are known to enforce each other (Hashish, 2019; Flaherty *et al*., 2019; Xu *et al*., 2020). This multifaceted aspect presents fundamental implications for career guidance where specific aspects of students' professional identity can be targeted and that an improvement in one construct could assist in improving another construct. It is therefore important to comprehend a global view of professional identity, however, construct interplay has mainly been examined for non-engineering students in studies that only focus on a restricted set of constructs. For instance, Fouad (1995) found that career awareness could be increased by actively promoting career exploration in American high school students. From the perspective of career guidance interventions, these literary limitations stress the need for a more general understanding of the construct interplay, especially in engineering students. This study will therefore focus on the interplay between multiple identity constructs, including the aforementioned constructs of *career exploration*, *self-awareness*, *career awareness*, *career-fit confidence*[1] and *competence confidence*, among engineering students.

1.3 Implications of professional identity development: why is it important?

The positive impact from the development of a professional identity has shown its relevance for non-engineering populations in both an educational and professional setting. During education, a positive association has been observed between students' self-believe regarding their academic skills and their academic achievements (Jaiswal & Choudhuri, 2017). Additionally, the academic success seems to be hampered when students focus strongly on their academic scores in the absence of a connection between their professional goals and educational track (Schneider *et al*., 1999). Both findings signify the importance of self-exploration and career exploration, even in the educational setting. Moreover, a larger correspondence between students' professional interests and their academic major also increased their satisfaction for their education (Fu *et al*., 2019). Consequently, a higher educational satisfaction has been shown to positively affect students' academic achievement (Balkis, 2013). Both the academic satisfaction and achievement have been shown to be guided by student's believe in their competences, as well as their expectations regarding the programme (Doménech-Betoret *et al*., 2017), again, indicating the importance of student's confidence and awareness during their educational career. In engineering students, it has been noted that engineering interest and competencies are positively related to the development of an engineering identity, which is linked to educational persistence in the engineering setting as well (Choe & Borrego, 2019), also indicating the importance of identity development in the explicit engineering population.

Similar to the educational setting, professional identity development has also been linked to career prospects in the professional setting. Like in education, compliance of one's interests with the pursued career is associated with higher job satisfaction and higher professional success (Bretz & Judge, 1994; Carson & Mowsesian, 1993). It has been noted that assessing one's competences drives specific career aspirations (Correll, 2004), and that it may also result in a longer employment in the first job (Cherry, 1974). These findings support the importance of developing self-awareness for the future career of students. Additional identity implications have for instance been noted for career awareness, as people with more

[1] In this study, the construct of *career-fit confidence* is conceptualized with respect to a professional engineering role as developed by Craps *et al*. (2021). Therefore, this construct is here denoted as *role-fit confidence* instead of *career-fit confidence* as employed by Cech *et al*. (2011).

knowledge regarding the professional world seem to make more successful choices (Shevlin *et al*., 2006). Self-awareness and career exploration have both been linked to an increased confidence in career decision-making (El-Hassan & Ghalayini, 2019; Özek & Ferraris, 2018). Career-fit confidence has been linked to the persistence in engineering (Cech *et al*., 2011), and the confidence in one's abilities has been linked to an improved work performance, active engagement, and higher persistence upon obstacles (Betz, 2000). Additionally, motivated engineers showed to stay longer in their job and are more productive (Beecham *et al*., 2008). These professional implications highlight the importance for students to develop a professional identity, and for education to provide guidance to students that require it. Therefore, a good understanding of the professional identity of engineering students is desirable.

1.4 Promoting professional identity development through career guidance

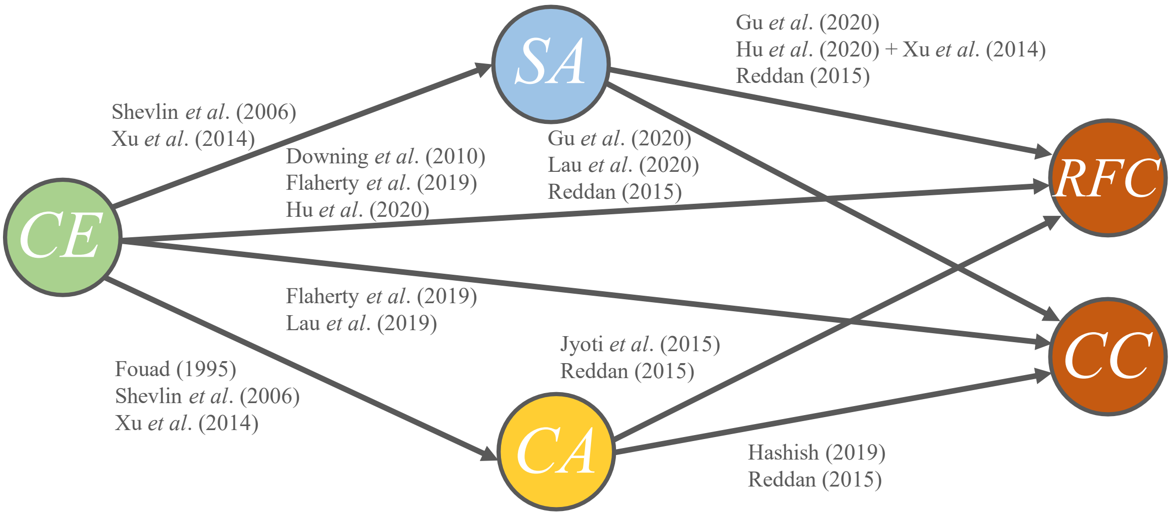
It was already postulated around the 1950s that an underdeveloped professional identity hampers the school to work transition (Vondracek, 1992). Since then, the advantages of a developed identity have been the driving force for providing students with applied professional knowledge to help their professional transition (Dik *et al*., 2011; Watts, 2006). As a result, different guidance programmes have been developed and implemented in curricula globally over the years, with the dual purpose of providing general professional knowledge and promoting personal development by stimulating one or multiple constructs that underly identity development (Gu *et al*., 2020; Keumala *et al*., 2018; Solberg *et al*., 2002; van Eeghen *et al*., 2019; Watts, 1997). These interventions have proven to be successful in helping students obtain the necessary insights and they have shown their efficacy in improving various identity aspect of students, including self-awareness, career awareness, decision-making, readiness, confidence, affinity with the field, social and professional abilities, academic achievement, or self-esteem (Chin *et al*., 2020; Flaherty *et* al., 2019; Gu *et al*., 2020; Lent *et al*., 2019; Lie *et al*., 2013; Reddan, 2015).

Intervention programmes have also provided fundamental evidence for the complex interplay between identity constructs by promoting a specific identity construct through the stimulation of another. However, a limitation of the current identity literature is that these intervention programmes have hardly been validated in engineering education and therefore do not provide evidence of such construct interactions for these students (Morelock, 2017; Figure 1). Recent advances were made by the KU Leuven Faculty of Engineering Technology to address this problem by designing a career guidance programme for engineering students that should be operationalized further in future education (Langie & Craps, 2020). Additionally, the intervention studies in literature have been performed in separate cohorts of different student populations across the world (Figure 1). The current literature thereby lacks a comprehensive study that examines the interplay between various constructs simultaneously within the same cohort, especially for engineering students, which would be of substantial value for the development of dedicated career guidance. This limitation is addressed in this study by hypothesizing a model for the interplay between five identity constructs, which could subsequently be used to investigate the construct interplay in the setting of engineering students from the Faculty of Engineering Technology at KU Leuven. The included constructs in the model are *career exploration*, *self-awareness*, *career awareness*, *role-fit confidence*, and *competence confidence*, which have not been investigated together in previous studies. The hypothesized model is presented in Figure 1 and is based on the available non-engineering intervention literature. It should be noted that the identity literature is often inconsistent regarding construct terminology or the construct representation and that not all constructs have been subjected to intervention studies so far. Therefore, effects among the constructs are hypothesized based on literary constructs with a close interpretation to the construct of interest (Appendix). For instance, Fouad (1995) showed that the 'occupational knowledge' of high school students increased after a career exploration intervention, which indicates a potential directional effect of *career exploration* on *career awareness*. Another example is presented by the study of Gu *et al*. (2020) on Chinese high school students. An intervention was provided that stimulated the understanding of their 'self-portrait', which was defined based on their interests, personality and abilities, which, although not explicitly stated, is representative for *self-awareness*. The intervention improved the 'career-decision self-efficacy' which represented the confidence in their skills such as planning and problem solving and the confidence in goal selection. From these effects, a direct link between self-awareness and confidence might be hypothesized.

Based on these example studies and those presented in Figure 1 and the Appendix, various potential directional effects between identity constructs could be included in the path model. These allow to hypothesize that complex direct effects could be present between identity constructs in engineering students, and that constructs might be affected in a way that positive construct stimulation originates at career exploration and subsequently affects awareness and confidence. These hypotheses are tested in this study on the same cohort of engineering students to obtain a more global understanding of their professional identity.

1.5 Professional identity development is affected by personal variables

In addition to the interplay between identity constructs, a second aspect of the complexity of professional identity is highlighted by the multiple personal variables that have been associated with its development. The presence of such influential factors has raised suggestions to provide personalized career guidance to students to comply with their specific needs (Hsieh & Huang, 2014; Ochs & Roessler, 2004; Taveira *et al*. 1998). Although of fundamental importance, such research is again mainly restricted to non-engineering students, providing an area for additional research in engineering education (Patrick &



Study	Student population	Employed identity aspect suggestive for				
		career exploration	*self-awareness*	*career awareness*	*role-fit confidence*	*competence confidence*
Downing *et al*. (2010)	American university students (n=285)	'career exploration'			'inability to specify a career goal'	
Flaherty *et al*. (2019)	American science university students (n=37)	'conference attendance'			'sense of belonging to the field'	'professional and social abilities'
Fouad (1995)	American early high school students (n=118)	'career exploration'		'occupational knowledge'		
Gu *et al*. (2020)	Chinese high school students (n=413)		'self-portrait'		'career decision self-efficacy'	'career decision self-efficacy'
Hashish (2019)	Egyptian nursing university students (n=245)			'career awareness'		'career and talent development self-efficacy'
Hu *et al*. (2020)	Chinese STEM university students (n=286)	'career exploration'	'self-exploration'		'perceived person-job fit'	
Lau *et al*. (2019)	Malaysian high school students (n=139)	'career exploration'	'self-awareness'			'career maturity'
Lau et al. (2020)	Malaysian vocational students (n=574)		'self-concept'			'work readiness'
Reddan (2015)	Australian science university students (n=15)		'self-awareness'	'opportunity awareness'	'goal selection'	'planning and problem solving'
Shevlin *et al.* (2006)	Irish high school students (n=325)	suggestive statement°	suggestive statement°	suggestive statement°		
Xu *et al*. (2014)	Chinese mixed university students (n=911)*	'environmental exploration'	'self-exploration' & 'lack of information about self'	'lack of information about occupations'		

° The paper states a suggestive effect among identity aspects instead of performing their own intervention study.
* 45 % of the students were from a STEM major.

Figure 1. Hypothesized path model for the identity constructs. The model was based on available literature, which constituted non-engineering research populations. CE: career exploration, SA: self-awareness, CA: career awareness, RFC: role-fit confidence, CC: competence confidence. See Appendix for a description of the corresponding literature.

Borrego, 2016; Morelock, 2017). Throughout previous studies, factors that have been considered key in identity development are the parents and the social environment. They are known to improve career exploration, decision-making, self-awareness, career awareness and confidence through a process of emotional support and interactive knowledge sharing (Bennett *et al*., 2016; Chin *et al*., 2020; El-Hassan & Ghalayini, 2019; Kracke, 2002; Rogers *et al*., 2008; Stringer & Kerpelman, 2010). Gender has also been included often in identity research, demonstrating that females perceive more stress during the career orientation process, and that they show more self-exploration, higher academic self-awareness but lower confidence (Correll, 2004; Jaiswal & Choudhuri, 2017; Nauta, 2007; Taveira *et al*., 1998; van Veelen *et al*., 2019). The relationship with other variables has been examined as well, for instance for socioeconomic status, personality and academic achievement (Bennett *et al*., 2016; Hsieh & Huang, 2014; Hu *et al*., 2020; Rogers *et al*., 2008). These studies mainly examined the relationship of a restricted number of personal variables on one or a few specific identity constructs. This, together with a low focus on engineering students, demonstrates the need for a larger investigation of influential factors. The present work will therefore examine the effect of multiple personal factors on the professional identity of engineering students.

1.6 Goals of the current study

The presented work aims to explore the broad complexity of the professional identity of engineering students, thereby adding to the fundamental insights for this research population that has not been investigated extensively so far. This is achieved by examining multiple latent identity constructs simultaneously, which is an approach that is currently lacking in literature. The included constructs are *career exploration*, *self-awareness*, *career awareness*, *role-fit confidence*, and *competence confidence*, which are each measured by dedicated sets of survey questions. Using the survey data and structural equation modeling techniques, several objectives were addressed in this work. First, the survey design was validated for measuring these identity constructs using an exploratory factor analysis. Second, the association structure between the constructs was determined using a confirmatory factor analysis. Third, the construct interplay was examined in more detail by determining direct construct effects using the previously hypothesized path model based on the literature. Fourth, personal variables were identified for which the professional identity differs between students by means of structural equation modeling. This work thereby further contributes to the general view of the professional identity of engineering students by providing an extensive exploration in the same engineering cohort. The insights from this study are supportive for the development of future career guidance programmes that aim to assist in the professional development of engineering students.

2 Instruments and methods

2.1 Instruments: participants and survey

Participants

Survey data was collected from engineering students at the Faculty of Engineering Technology at KU Leuven. Students participated in either May 2019 or 2020, resulting in 1040 entries in total and the data from these past two academic years was made available for this research. Participation was allowed for students from all phases in the programme, including students from the first, second and third bachelor year, the master year, and the transfer programme for graduate students from technical University Colleges prior to the master year.

The survey

The predeveloped cross-sectional survey was provided electronically via Qualitrics and lasted around 10 minutes. Ethical approval was granted from the university's Ethics Committee (G-2019 03 1596) and participants have consented to be part of the study after being informed about the voluntary nature of their participation and the anonymous data analysis.

In a first part of the survey, information on several background variables was collected, including professional role interest, parental occupation, and engineering persistence. A question was presented that probed the student's professional role interest where they had to indicate their preference for one of the professional engineering roles as described by Craps *et al*. (2021), which included product leadership (PL, focusing on innovation), operational excellence (OE, focusing on process and product optimization), customer intimacy (CI, focusing on client tailored solutions), their combinations (PL+OE, PL+CI, OE+CI, PL+OE+CI), or an option indicating that the student was still unaware of his/her preference. Two additional questions asked whether one of the student's parents was an engineer or not, which is referred to as parental occupation, and whether the student was considering another job outside engineering with responses 'yes', 'no', or 'sometimes', referred to as engineering persistence. Additional information regarding the student's gender, phase of study and migration status was provided by the university anonymously.

In a second part of the survey, the student's attitude was probed towards five putative professional identity constructs, which are each measured by a set of designed Likert scale questions (Table 1). The five constructs included *career exploration*, *career awareness*, *self-awareness*, *role-fit confidence*, and *competence confidence*. Based on the survey items assigned to the constructs, the interpretations of the constructs were as follows: (1) *career exploration* probed the extent of different information seeking behaviors, (2) *career awareness* probed the understanding of the professional engineering roles and their associated competencies following Craps *et al.* (2021), (3) *self-awareness* probed how well they envision themselves as an engineer along with their desires and abilities in the engineering field, (4) *role-fit confidence* probed the confidence towards their desired engineering role, and (5) *competence confidence* probed the confidence in their non-technical skills to succeed in that role. The survey questions for the two confidence constructs were based on Cech *et al.* (2011).

Table 1. Survey items per putative construct. The abbreviation for each question is shown next to the question along with the Likert scale. The number for each question reflects the order in the survey. The dots on the Likert scales indicate the number of response categories. Only the first and last response category are indicated for visibility. Intermediate dots correspond to intermediate ordinal categories. Response categories ‘strongly disagree’ and ‘strongly agree’ are here represented by ‘disagree’ and ‘agree’ for visibility. The survey was developed two years ago by Sofie Craps, the mentor of this thesis.

Career exploration		
“Do you sometimes participate in activities that are not associated to the educational programme, but which you consider contributing to becoming a good engineer, for example a (voluntary) company visit, a talk, networking evening, …?”	Q24	○—○—○—○ Never … Frequently
“I explore possible careers through reading, videos, or by talking to engineers”	Q25	○—○—○—○—○ Never … A lot
“I keep in touch with engineers and/or companies that might be relevant in the future”	Q26	○—○—○—○—○ Never … A lot
“I look for engineering job offers”	Q27	○—○—○—○—○ Never … A lot
“I go to job fairs or events with companies”	Q28	○—○—○—○—○ Never … A lot
“I talk to my parents, friends, lecturers or professionals about my opportunities on the labour market”	Q29	○—○—○—○—○ Never … A lot
“I try to explore and develop my strengths and weaknesses in order for me to find a job that suits me”	Q30	○—○—○—○—○ Never … A lot
“I have relevant work experience”	Q31	○—○—○—○—○ Never … A lot
Self-awareness		
“Can you see yourself as an engineer?”	Q1	○—○—○ No … Yes
“How do you picture yourself as engineer?”	Q2	○—○—○—○ Not … Clearly
“Do you find it difficult to position yourself in the roles model?”	Q7	○—○—○ No … Yes
“I know for a long time that this is the role I want to fulfil”	Q12	○—○—○—○ Disagree … Agree
“Would you choose the same professional role based on the competence profiles when thinking about your first ideal job?”	Q18	○—○ No … Yes
“Do you find it difficult to select a preferred professional role model based on the competencies?”	Q19	○—○—○ No … Yes

"I know what non-technical competencies I am good/not good at"	Q23	○—○—○—○ Disagree / Agree
Career awareness		
"I understand the description of the different professional roles"	Q3	○—○—○—○ Disagree / Agree
"I can relate the roles to the learning experiences in the educational programme"	Q4	○—○—○—○ Disagree / Agree
"I can link the roles to types of engineers in the field"	Q5	○—○—○—○ Disagree / Agree
"This model is not applicable to the engineers of my specific educational track"	Q6	○—○—○—○ Disagree / Agree
"I recognize most of the competencies in this model"	Q13	○—○—○—○ Disagree / Agree
"I understand the professional roles better based on the competencies"	Q14	○—○—○—○ Disagree / Agree
"I am not sure what to make of these competencies"	Q15	○—○—○—○ Disagree / Agree
"I don't think that all these competencies are important for engineers"	Q16	○—○—○—○ Disagree / Agree
"I don't think that an engineer can be successful in so many competencies"	Q17	○—○—○—○ Disagree / Agree
Role-fit confidence		
"This role is the suitable professional role for me"	Q8	○—○—○—○ Disagree / Agree
"A job in this professional role will give me a great work satisfaction"	Q9	○—○—○—○ Disagree / Agree
"I am trying my best to succeed in this role"	Q10	○—○—○—○ Disagree / Agree
"I am convinced that I chose the role that suits me best"	Q11	○—○—○—○ Disagree / Agree
Competence confidence		
"I possess the required non-technical competencies to develop my preferred engineering role"	Q20	○—○—○—○ Disagree / Agree
"I have sufficient insight in my talents and competencies to succeed in my preferred role"	Q21	○—○—○—○ Disagree / Agree
"I am good in non-technical competencies compared to my peers"	Q22	○—○—○—○ Disagree / Agree

2.2 Statistical analysis

Several analyses were performed to investigate the professional identity of engineering students. First, the specific constitution of each construct was determined in an exploratory phase to examine whether the designed survey supports the five putative constructs. This included a Horn's parallel analysis to explore the number of factors underlying all survey questions, and an exploratory factor analysis (EFA) to determine a putative factor model where the survey items are assigned to a specific construct. Second, in a confirmatory phase, the obtained model is explicitly tested using confirmatory factor analysis (CFA) and the resulting factor model is used to further investigate the constructs in a structural equation model (SEM). In this final stage, construct differences are determined for multiple background variables, i.e. gender, phase of study, parental occupation, professional role interest, engineering persistence, and migration status, as well as the investigation of the direct effects of constructs on each other.

Exploratory phase

First, the number of constructs represented by the survey was determined by a Horn's parallel analysis (Horn, 1965), which presents an estimation based on the eigenvalues of the survey item correlations and the eigenvalues of randomly simulated correlation matrices. 1500 iterations were used to create the random simulated data and both a factor analysis and principal component method were used to determine the eigenvalues, with the unweighted least squares (ULS) estimator. Despite the categorical nature of the data, pairwise Pearson correlations were used instead of polychoric correlations for categorical data to achieve convergence of the algorithm (Timmerman & Lorenzo-Seva, 2011). In addition to Horn's parallel analysis, the number of putative constructs was also explored by a visual examination of the polychoric-based eigenvalue pattern in a scree plot. This was guided by the elbow rule which proposes that the number of relevant constructs corresponds with the position at which a drop in the eigenvalues occurs (Sutono, 2016).

Second, the suggested number of constructs was implemented in the EFA to determine which survey items were associated with each construct and to provide construct interpretations. Additionally, the EFA allowed to identify survey items that did not fit well in the model, i.e. of which the factor loading did not reach a magnitude of 0.30 (Brown, 2006, p.130; Özek & Ferraris, 2018; van Veelen *et al*., 2017). The main factor on which an item loaded was determined by its strongest loading, and the discrepancy with other constructs was reached when the strongest loading had a magnitude difference of minimally 0.10 with the cross-loadings (Woo *et al*., 2018). Pairwise polychoric correlations were used, along with the oblique Promax rotation to allow correlation between the constructs. Although the diagonally weighted least squares (DWLS) estimator has been proposed for the analysis of categorical data, this estimator was not available for the EFA, nor for Horn's parallel analysis, thereby opting for the use of the ULS estimator because of its similar performance as the DWLS estimator for ordinal categorical data (Forero *et al*., 2009).

The exploratory phase was performed on a random 40 % subset of the data (n=416) to serve as a training set for the development of the factor model for the constructs. This random training set did not differ from the complementary 60 % subset (Suppl. Table 1) based on chi-squared tests.

Confirmatory phase

The factor model obtained in the exploratory phase was validated in a CFA and was implemented in a SEM using the complementary random 60 % subset (n=624). The CFA and SEM analyses were performed on 20 multiply imputed datasets to account for the missingness in the survey data.

First, an ordinal categorical CFA (Muthén, 1984) was performed using polychoric correlations and the DWLS estimator (Li, 2016a; Svetina *et al*., 2019). The CFA model was validated on multiple levels, i.e. the global fit by looking at the averaged model fit indices, the local fit by examining the modification indices and whether the magnitude of the factor loadings exceed 0.30. For the global fit, four model fit indices were reported, including the comparative fit index (CFI), Tucker-Lewis index (TLI), root mean squared error of approximation (RMSEA), and the standardized root mean squared residual (SRMR). Values exceeding 0.90 for the CFI and TLI, and values below 0.08 for the RMSEA and SRMR signify appropriate model fit (Brown, 2006, p.87; Hooper *et al*., 2008; Laverdière *et al*., 2013; Shevlin & Millar, 2006). For the local fit, modification indices larger than 20 indicate improper fit. Additionally, appropriate discrimination between the constructs was obtained when the corresponding factor correlation fell below 0.80 (Brown, 2006, p.131).

Second, two SEM analyses were performed that, first, determined construct differences for multiple background variables, referred to as a multiple indicator multiple cause (MIMIC) model, and second, a SEM analysis to examine directional effects between constructs in a path model. The CFA model was used in both analyses as measurement model for the constructs, including polychoric correlations and the DWLS estimator. In the first SEM analysis, the categorical background variables were incorporated in the structural model part of the MIMIC model as dummy variables to examine the difference in construct level between a category and the variable's reference group. In the second analysis, the structural model consisted of a directional path model between the constructs, which was based on the previous literature (Figure 1). The reported factor loadings in the CFA and SEM analysis were obtained using standardized factors only and are denoted as standardized factor loadings (Brown, 2006, p.137; Laverdière *et al*., 2013).

Multiple imputation

Data imputation was based on regression models that included the background variables, and survey items that showed an absolute correlation of minimally 0.25 with the imputed item. For ordinal survey items and background variables, an ordinal logistic regression was used for imputation, while for nominal survey items and background variables polytomous logistic regression was employed. A set of 20 imputed datasets was generated from 20 iterations each.

Details

Analyses were performed in Rstudio (R Core Team, 2019). Throughout the study, significance was reached when the p-value fell below 0.05 and significance levels were indicated according to not significant (n.s.) ≥ 0.05 > * ≥ 0.10 > ** ≥ 0.001 > ***. P-values were adjusted by the false discovery rate (FDR) when multiple comparisons were performed. All random algorithms were executed with a seed of 1234. Horn's parallel analysis and EFA were performed using the psych package (Revelle, 2014), multiple impution based CFA and SEM were performed with the semTools package (Jorgensen *et al*., 2018), and multiple imputation itself was performed using the mice package (van Buuren & Groothuis-Oudshoorn, 2011). Throughout the report, figures and tables are presented with hyperlinks to direct the reader to the corresponding information. For each result section, R code is made available on Github which can be accessed via the links in the Appendix.

3. Results

3.1 Descriptive analysis

3.1.1. Characterization and assessment of the survey items and background variables

Survey items

Prior to analyzing the survey data, a descriptive analysis is presented to characterize the survey items and to assess their applicability for factor analysis. The 31 questions probing career attitudes are provided chronologically in Table 2. Response frequencies per Likert point are presented in the heatmap, showing that low frequencies are mostly associated with boundary Likert points and that the central Likert points are mainly preferred by the students. No extreme response behavior is present where one Likert point was consistently favored over the others by all students, which would impede subsequent factor analysis (Lorenzo-Seva & Ferrando, 2020). The number of responses per question varies as the incidence of missing values increased along the survey, reaching almost 16 % missingness at the final question.

The quality of the survey questions for factor analysis was examined by the Kaiser-Meyer-Olkin (KMO) values and polychoric correlations. Two questions showed low KMO values between 0.50 and 0.60 (Q2 and Q14) and three questions showed values between 0.60 and 0.70 (Q16, Q17 and Q23; Table 2), suggesting a possible unsatisfactory association of these questions with the other survey items (Kaiser & Rice, 1974). Indeed, the polychoric correlation matrix shows that these questions lack substantial correlations with other items (Suppl. Figure 1). In contrast, the other questions show several high inter-item correlations resulting in several item clusters that could be representative for the identity constructs. For instance, the final eight items ranging from Q24 to Q31 are substantially positively correlated with each other and show only a limited number of cross-correlations with other items. This signifies a distinct item subset where an increased response level in one question is associated with an increased level in the other items. The presence of such correlation groups supports the use of factor analysis and structural equation modelling techniques since these rely on inter-item correlations.

Background variables

Throughout the study, six background variables were included in the analyses, i.e. gender, phase of study, parental occupation, professional role interest, migration status, and engineering persistence (Table 3). Participants were predominantly male, analogously to the registered student population at the Faculty of Engineering Technology (85.16 % vs. 84.90 %). First year bachelor students constituted the largest group of respondents, followed by the other bachelor years. Proportions for the study phases were representative for the registered student population (32.10 % vs. 29.75 %, 23.86 % vs. 19.79 %, 23.35 % vs. 22.65 %, 6.52 % vs. 6.70 %, 15.18 % vs. 21.11 %). Most of the engineering students came

Table 2. Descriptive statistics of survey questions. Items are ordered as they appeared in the survey. Absolute counts per Likert responses are represented in the heatmap. Empty cells are not part of the Likert scale. Total number of responses per question equals the sum of the absolute counts. Percentage of missingness was calculated on the total number of respondents (n=1040). KMO values are reported per survey item.

Survey item	Likert responses					Total	Missing (%)	KMO
	1	2	3	4	5			
Q1	58	617	365			1040	0.00	0.86
Q2	58	268	429	264		1019	2.02	0.53
Q3	11	25	564	378		978	5.96	0.79
Q4	29	255	624	70		978	5.96	0.73
Q5	18	142	659	158		977	6.06	0.84
Q6	232	620	105	20		977	6.06	0.81
Q7	302	424	244			970	6.73	0.81
Q8	9	118	701	102		930	10.58	0.83
Q9	10	45	584	292		931	10.48	0.87
Q10	10	62	571	288		931	10.48	0.89
Q11	48	340	446	96		930	10.58	0.86
Q12	197	469	216	49		931	10.48	0.85
Q13	1	41	661	207		910	12.50	0.79
Q14	8	170	554	178		910	12.50	0.52
Q15	215	609	77	9		910	12.50	0.83
Q16	226	490	164	30		910	12.50	0.68
Q17	142	495	246	27		910	12.50	0.64
Q18	200	711				911	12.40	0.76
Q19	342	381	184			907	12.79	0.76
Q20	18	278	527	69		892	14.23	0.81
Q21	8	179	622	82		891	14.33	0.87
Q22	17	273	453	148		891	14.33	0.86
Q23	26	179	569	118		892	14.23	0.63
Q24	162	370	257	100		889	14.52	0.83
Q25	74	263	299	196	54	886	14.81	0.89
Q26	272	288	190	92	32	874	15.96	0.90
Q27	271	217	192	132	65	877	15.67	0.86
Q28	348	257	167	73	27	872	16.15	0.79
Q29	76	175	304	247	80	882	15.19	0.88
Q30	63	127	296	310	86	882	15.19	0.90
Q31	313	304	154	79	26	876	15.77	0.90

from a household where the parents were not engineers and most students did not have a migration background. The minority of engineering students were considering another job outside engineering, while similar frequencies were reached for students who reported to persist in engineering and those who were sometimes reconsidering it. Finally, most students were able to indicate their desired future professional engineering role. The ability of students to identify with a professional role was independent of their engineering persistence ($p=0.92$, $\chi^2=0.175$, $df=2$).

Table 3. Descriptive statistics of background variables. The absolute (count) and relative (%) frequency are presented per category for each variable. Relative frequencies were calculated on non-missing data.

	Category	Frequency		Missing (%)
		Count	(%)	
Gender	Male	878	(85.16)	0.87
	Female	153	(14.84)	
Phase of study	1st bachelor	330	(32.10)	1.15
	2nd bachelor	235	(23.86)	
	3rd bachelor	240	(23.35)	
	Transfer student	67	(6.52)	
	Master student	156	(15.18)	
Parental occumation	At least one engineering parent	295	(28.53)	0.58
	Non-engineering parents	739	(71.47)	
Vocational interest	No idea	93	(9.59)	6.73
	PL	153	(15.77)	
	CI	78	(8.04)	
	OE	85	(8.76)	
	PL+OE+CI	66	(6.80)	
	PL+OE	179	(18.45)	
	PL+CI	165	(17.01)	
	OE+CI	151	(15.57)	
Migration status	Migration background	70	(7.20)	6.54
	Non-migration background	885	(91.05)	
	Other	17	(1.75)	
Engineering persistence	Persistent	331	(37.23)	14.52
	Sometimes thinking about a non-engineering career	414	(46.57)	
	Thinking about a non-engineering career	144	(16.20)	

3.1.2. Investigating missingness

Missingness throughout the survey is examined because it can potentially bias subsequent results. Therefore, the present missingness patterns are explored along with the potential missingness mechanism. For the latter, the influence of several background variables and survey responses on the missingness is determined.

Missingness patterns

Table 2 showed an increase in missingness throughout the survey, suggesting a systematic non-response by the students while progressing the survey. The majority of the students completed the survey without skipping any questions, resulting in 859 completers (82.60 %) (Figure 2A). Many of the 181 non-completers showed large missingness with over 18 missing values, corresponding with 60 % of the survey. Among the 181 non-completers, 32 different missingness patterns existed (Figure 2B). Of these, only 28 resembled an intermitted missingness pattern while 153 resembled a monotone pattern which indicated

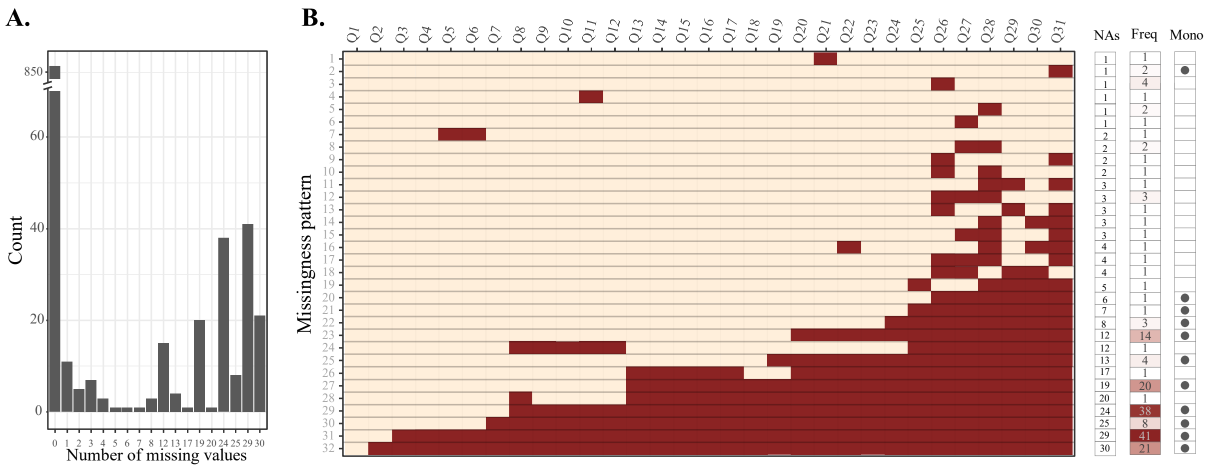


Figure 2. A. Frequency of number of missing values per student. Remark the break in the y-axis. **B.** Missingness patterns. Red: missingness, yellow: observed. Rows represent the 32 unique missingness patterns from 181 non-completers. Complete data was observed for 859 cases (82.60 %). Patterns are ordered from top to bottom by increasing number of missingness. On the right, the number of missing values in the pattern is shown (NAs), along with the occurrence frequency for each pattern (Freq) and whether the missingness pattern in monotone (dot, n=153) or intermitted (no dot, n=28) (Mono). 129 monotone dropout patterns started before Q16. Survey items at the top are ordered chronologically.

that the student had quit the survey. Such dropout occurred most frequently in the first half of the survey as 85 % of the monotone dropout patterns started before Q16.

Missingness mechanism

It is important to examine whether missingness is related with the survey responses, and hence, possibly with the professional identity. Therefore, the effect of the obtained data on the probability of dropping out is examined following two approaches, i.e. multiple comparisons (Karbownik *et al.*, 2021) and logistic regression (Verbeke *et al.*, 2015, p.352). Since missingness predominantly occurred as dropout from the survey, only measurements before the first missing value can be used to assess their dependency on subsequent dropout. Additionally, the effect on dropout cannot be examined for individual survey items because of the low sample size per missingness pattern. Considering these two limitations, dropout was investigated for five consecutive parts of the survey to increase the sample size (Suppl. Figure 2A). Dropout was then investigated for participants that dropped out in the respective survey part with respect to students that had not yet dropped out. In the analysis, the dropout time was determined as the question at which the first missingness event occurred.

As a first assessment, Fisher exact tests were used to examine the association between dropout in a respective survey part and the obtained data. Results were corrected for multiple comparison using the false discovery rate. The background variables gender, phase and parents showed no significant effects on missingness across the five parts of the survey.

Only four significant associations were present between survey responses and subsequent missingness. Dropout in part 4 and 5 seemed to be affected by the response behavior at Q2 and Q9, and Q6 and Q9, respectively. However, these pairwise comparisons are only explorative as only a single variable is considered. Therefore, in a next step, regression methods are employed to investigate the missingness by including multiple variables simultaneously.

Logistic regressions were constructed to determine the effect of the data on the probability to complete the five survey parts. However, variables often had a zero count for one or more response categories for dropout groups in parts 3, 4 and 5 (Suppl. Figure 2B), resulting in data separation induced model non-convergence. Given that the larger completers group also showed low abundances for these response categories, it is reasonable to assume that the zero counts arose because of low sample sizes in the dropout groups. This probabilistic data separation issue can be resolved with the Firth model correction, however, this did not resolve the issue for the models for part 4 and 5. Therefore, parts 3, 4 and 5 were combined into one model. The model results show that the probability to complete the first survey part is affected by the phase of study and Q1, whereas completing the second part was only affected by Q1 and the third part only by Q4 (Suppl. Table 2). Although these results are only explorative because of the manifested issues, they suggest a possible dependency between dropping out of the survey and the recorded data. Therefore, in the confirmatory phase of this study, multiple imputation will be applied to restore the missing values based on the response behavior and background information of the students.

3.2 Exploratory analyses suggested five survey constructs

A first objective of this study was to examine the appropriateness of the survey in measuring the hypothesized identity constructs. This is investigated via a sequential approach: first, the putative number of constructs present in the survey is determined by Horn's parallel analysis, and second, the representation of the constructs is determined with an exploratory factor analysis.

Horn's parallel analysis

The putative number of constructs captured by the survey is examined via the eigenvalue pattern of the correlation matrix of the survey items. First, a Horn's parallel analysis based on both a factor analysis and principal component analysis suggested 9 factors and 5 components respectively, as determined by the crossing of the scree plot for the actual data and both the simulated and resampled data (Figure 3A). Second, eigenvalues based on polychoric correlations also suggested a plausible range of factors or components (Figure 3B). Based on Kaiser's criterion arguing for the retention of eigenvalues larger than one, 8 or 9 factors/components could be identified. Additionally, 5 or 9 factors/components could be

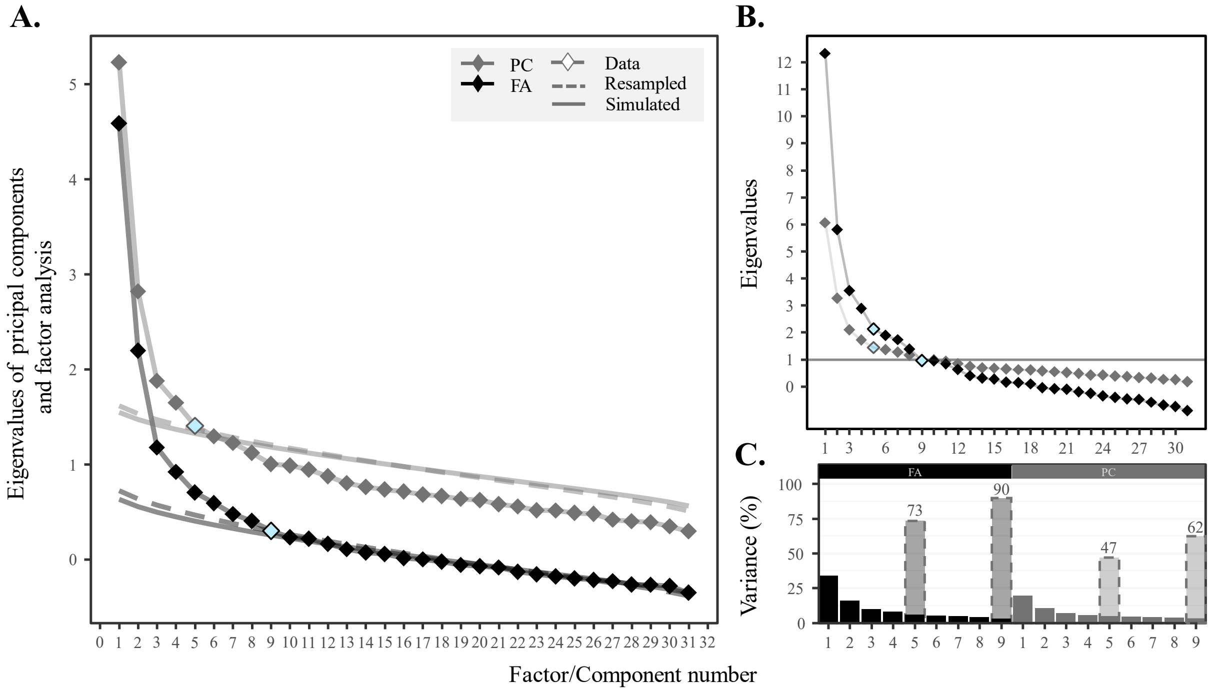


Figure 3. Horn's parallel analysis. FA: factor analysis, PC: principal component analysis. **A.** Parallel analysis based on Pearson correlations. Analysis did not converge with polychoric correlations. Blue dots indicate the optimal number of factors/components. **B.** Scree plot of eigenvalues based on polychoric correlations for PC (gray) and FA (black). **C.** Variance per factor/component (plain) and total variance for five and nine factors/components (dashed). Variances were calculated with the positive polychoric based eigenvalues.

extracted according to the elbow rule which suggests that a possible nick in the eigenvalue pattern corresponds with the optimal number of factors. Based on these three criteria, a possible range from 5 to 9 constructs could be supported by the survey, which together comprise a substantial part of the variance in the data (Figure 3C).

Exploratory factor analysis (EFA)

Given the putative number of constructs, an EFA is performed to unravel the construct interpretations according to the specific items that they support, as well as to identify suboptimal survey items that are not represented well in the analysis. As suggested by the parallel analysis, nine factors were included in an initial EFA. Factor loadings revealed that two of the nine constructs were represented by only two items, that some items failed to load on any factor, and that multiple items showed substantial cross-loadings with other factors. This insufficient factor representation could result from the overestimation of the number of factors, leading to auxiliary constructs. Therefore, the number of factors was systematically reduced to obtain a factor solution with more than two items per factor, which was satisfied for a six-factor model. However, this model included three items with low loadings for all factors, i.e. Q22, Q14, and Q23, of which the last two had demonstrated low correlations with other variables before (Suppl. Figure 1), which is suggestive for their unsatisfactory behavior in the factor analysis. These redundant items were therefore sequentially removed, along with the malfitting items Q1, Q2 and Q18, resulting in a five-factor model with substantial factor loadings and minimally three items per factor (Table 4). This final model still included two substantial cross-loadings larger than 0.30, i.e. for Q3 and Q9. However, these were retained since the difference between the cross-loading and the main loading exceeded 0.10, as suggested by Woo *et al.* (2018).

The high-loading survey items probe a similar topic within each of the factors. This allows to develop five distinct interpretations for the emerged constructs. The first construct includes the eight original survey items that reflect the intensity of career exploration and is therefore referred to as *career exploration*. The second construct contains three questions that probe how well the students understand the different professional engineering roles, defining this construct as *professional roles awareness*. Similarly, four items in the third construct probe how well the students understand the competencies that are associated with the different professional roles, defining this construct as *competence awareness*. However, Q6, which

Table 4. Five-factor EFA model. Q6 was eventually removed from the construct because of low correspondence with the construct's interpretation. Variance proportions for the factors were 0.13 (F1), 0.09 (F2), 0.07 (F3), 0.06 (F4), 0.05 (F5).

	F1	F2	F3	F4	F5	Interpretation
Q24	**0.50**	0.03	-0.03	0.00	0.02	Career exploration
Q25	**0.75**	0.01	-0.02	0.02	-0.02	
Q26	**0.74**	0.03	0.02	-0.06	0.02	
Q27	**0.72**	-0.02	-0.01	0.05	-0.04	
Q28	**0.72**	-0.12	0.03	0.04	0.01	
Q29	**0.59**	0.03	-0.04	0.01	-0.06	
Q30	**0.50**	0.06	-0.17	-0.09	0.04	
Q31	**0.53**	-0.12	0.19	0.09	0.27	
Q3	-0.07	-0.10	**-0.32**	**0.66**	-0.05	Professional roles awareness
Q4	0.00	0.03	0.15	**0.59**	0.01	
Q5	0.10	-0.05	-0.09	**0.67**	-0.01	
~~Q6~~	~~0.01~~	~~-0.07~~	~~**0.45**~~	~~-0.15~~	~~0.09~~	Competence awareness
Q13	-0.01	0.04	**-0.39**	0.26	0.16	
Q15	0.07	0.04	**0.50**	-0.20	-0.28	
Q16	-0.05	0.14	**0.63**	0.03	-0.09	
Q17	0.01	0.00	**0.47**	0.09	-0.15	
Q7	0.11	**-0.54**	-0.04	-0.12	-0.14	Role-fit confidence
Q12	0.11	**0.52**	0.21	-0.02	0.25	
Q8	-0.11	**0.80**	-0.03	-0.20	0.11	
Q9	-0.01	**0.59**	**-0.41**	-0.04	-0.12	
Q10	0.20	**0.50**	-0.17	0.07	-0.25	
Q11	0.03	**0.67**	0.12	0.10	0.09	
Q19	0.07	-0.07	0.10	-0.09	**-0.53**	Competence confidence
Q20	0.01	0.06	-0.04	-0.08	**0.48**	
Q21	-0.03	0.20	-0.03	0.13	**0.36**	
Q22	0.08	-0.01	-0.26	-0.19	**0.51**	

was a fifth item that also showed a high loading for this construct, was removed from this construct because the topic of this item did not correspond with the interpretation of the construct. The fourth construct comprises six items that measure the student's belief of how well their chosen engineering role suits them, resulting in the construct *role-fit confidence*. Finally, the fifth construct is composed of four items reflecting their confidence in their non-technical competencies, defining this construct as *competence confidence*.

This EFA shows that the majority of the designed survey items load in the previously hypothesized identity construct. However, this does not apply for the items of *self-awareness*. Several of these items did not fit well in the factor model while three items loaded on the confidence constructs. Therefore, the designed *self-awareness* items were not effective in representing this hypothesized construct. Additionally, the items corresponding with the original *career awareness* construct loaded in two separate constructs, i.e. *professional roles awareness* and *competence awareness*, indicating that the hypothesized *career awareness* construct consists of two subdimensions.

3.3 Validation of the five-construct model

Confirmatory factor analysis (CFA): final model development

The retained model from the exploratory analysis is now explicitly tested in a CFA to see if it sufficiently fits the data. In the first step of the confirmatory approach, the CFA model configuration is specified. The five respective constructs in this model included only the high-loading items as reported by the EFA without including the cross-loadings, to model item specificity. Dedicated anchor items were chosen for each construct that demonstrated large factor loadings and low cross-loadings in the EFA and of which an increase on the item's response scale aligns with the interpretation of the construct. For instance, Q13 was depicted as the anchor item for *competence awareness* over the other three items since a higher response value for this item corresponds with an increased understanding of the competencies, corresponding with the intention of the construct.

In a second step, model identification is examined to ensure model convergence. This proposed CFA model configuration resulted in a desirable over-identified model where the input information exceeds the number of parameters to be estimated. Specifically, with the 25 observed survey items and the 60 parameters to estimate, the model retained 265 degrees of freedom (Calculation 1).

The third step involved the fitting of the model and evaluating its fit. All factor loadings reached statistical significance with magnitudes exceeding 0.30 (Suppl. Table 3A) and reasonable model fit indices (Model 1, Table 5). However, modification indices identified local misfit in several parts of the model. In total, seven model constraints were identified, of which

$$\begin{aligned}
&\textit{Inputs} = 25 * (25 + 1) / 2 = 325 \\
&(1)\ \textit{Factor correlations} = 5 * (5 - 1) / 2 = 10 \\
&(2)\ \textit{Factor loadings} = 25 \\
&(3)\ \textit{Error terms} = 25 \\
&\textit{Parameters} = (1) + (2) + (3) = 60 \\
&\text{Degrees of freedom} = \textit{Inputs} - \textit{Parameters} = 325 - 60 = 265
\end{aligned} \qquad (\textit{Calculation 1})$$

five suggested the introduction of residual correlation between survey items to improve model fit, and two others suggested the introduction of cross-loadings with other constructs (Suppl. Table 3B). To prevent overfitting of the model, only reasonable modifications were considered to improve the model. The suggested cross-loadings included items of which the meaning did not correspond with the respective construct and were therefore not implemented. Three of the suggested correlations corresponded to items that showed similar characteristics in content or wording, i.e. the item duos Q7 - Q19 and Q24 - Q28 probed the same topic of role identification difficulty and company exploration activities, respectively, while Q16 and Q17 were questions that were both phrased negatively. Such similar characteristics are known to cause additional correlation between items apart from the underlying latent construct, thereby justifying the implementation of residual correlation between these items to correct for correlation that is not attributed by the latent factor (Raykov, 2004). Allowing these three modifications improved the model significantly ($p<0.001$, $\chi^2=114.873$, df=3). Additionally, one suggested cross-loading was dropped by accounting for residual correlation in the model. The remaining suggested cross-loading was again not in line with the construct interpretation and the magnitudes of the corresponding factor loading only reached a trivial value (Suppl. Table 4B). Finally, the residual correlation between Q16 and Q17 caused their factor loadings to drop, with a remaining magnitude slightly above 0.30 (Suppl Table 3A & 4A), because a part of their previous correlation with the construct is now captured by the residual correlation.

Table 5. Model fit when adding residual correlation. Model 1 represents the initial model without residual correlations. The other models sequentially include an extra correlation term to the previous model. Model results for model 1 (initial model) and model 4 (final model) are presented in Suppl Tables 3 & 4. Results were obtained from 20 multiply imputed dataset.

	CFI	TLI	RMSEA	SRMR	χ^2 (df)	p-value
Model 1	0.955	0.949	0.057	0.065	457.944 (265)	0.000
Model 2 (+ Q7 ~~ Q19)	0.965	0.960	0.051	0.061	370.457 (264)	0.000
Model 3 (+ Q16 ~~ Q17)	0.969	0.965	0.048	0.059	367.766 (263)	0.000
Model 4 (+ Q24 ~~ Q28)	0.972	0.967	0.046	0.058	343.071 (262)	0.000

Final CFA model: model and construct validation

The developed model demonstrated proper model convergence in the structural equation analysis. No Heywood cases occurred where either variances were negative, correlations exceeded a value of one, or standard errors reached extreme values. Moreover, the input and model correlation matrices were not non-positive definite as all eigenvalues were larger than zero, further assuring the reliability of the converged results (Lorenzo-Seva *et al.*, 2020). Adequate model fit indices were obtained (Model 4, Table 5), with factor loadings reaching magnitudes above 0.30 of which the majority exceeded 0.50 (Figure 4, Suppl. Table 4A), indicating a sufficient factor representation. One item for *professional roles awareness* and *competence confidence* had a loading somewhat below 0.50. However, two questions for *competence awareness* reached minimal values slightly above the 0.30 threshold, suggesting a possible unstable behavior of these questions in the model.

In addition to the factor loadings, several characteristics of the constructs are examined, i.e. the Cronbach's alpha, the measure of sampling adequacy (MSA), and the averaged inter-item correlation (Table 6). First, Cronbach's alphas reached appropriate values for only two of the five constructs based on the general 0.70 threshold, namely for *career exploration* and *role-fit confidence*. The values for the other three constructs fell in the 0.50-0.60 interval, which assumes low construct reliability. Second, MSA values were again high for *career exploration* and *role-fit confidence*, reaching values above 0.80. The other three constructs had values in the 0.60-0.70 interval, representative of mediocre adequacy. Third, average inter-item correlations per construct indicated large values for *career exploration* and *role-fit confidence*, reasonable values for *professional roles awareness* and *competence confidence*, and the lowest value for *competence awareness*. Together, these three measures suggest an appropriate construct reliability and representation for *career exploration* and *career-fit confidence*, while those for the other three construct were less sufficient.

Table 6. Construct characteristics. Cronbach's alphas represent the standardized version. Polychoric correlations were used for the mean correlation.

	Career exploration	*Professional roles awareness*	*Comptence awareness*	*Role-fit confidence*	*Competence confidence*
Cronbach's alpha	0.83	0.54	0.51	0.77	0.57
MSA	0.86	0.62	0.60	0.80	0.67
Mean correlation	0.42	0.36	0.27	0.44	0.32

Final CFA model: model interpretation

The developed model presents a five-factor structure that includes *career exploration*, *professional roles awareness*, *competence awareness*, *role-fit confidence*, and *competence confidence*. The loadings show that, for example, all items for *career exploration* are positively related with the construct. An increase in *career exploration* therefore implies that students respond with larger values on the corresponding items. Similarly, negative loadings for *role-fit confidence*, *competence confidence*, and *competence awareness* show that an increase in these latent factors is associated with lower response values in the respective survey items. This is reasonable since, for instance, an increase in Q7 (*role-fit confidence*) represents a higher difficulty to identify with a professional role, which represents a lower confidence that a career might be suitable.

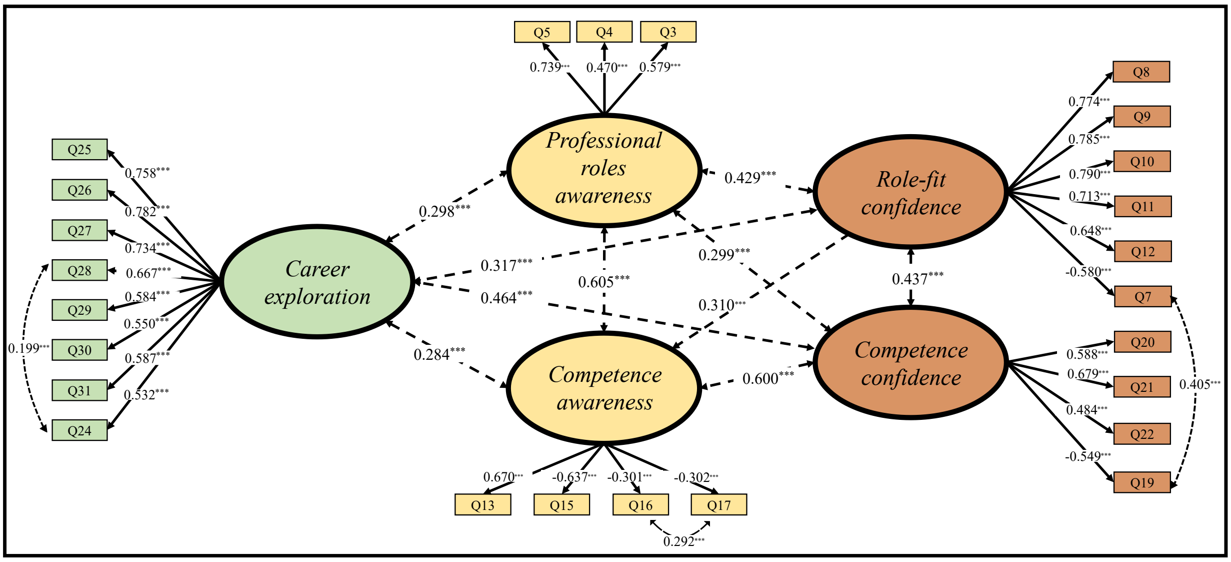


Figure 4. Final CFA model. The first item for each construct is the anchor item. Single-headed arrows represent standardized factor loadings for the survey items. Factor correlations and residual correlations are presented by the double-headed arrows between the constructs (ovals) and items (rectangles) respectively. Results are obtained from 20 multiply imputed datasets (n=624). Significance is indicated by asterisks (0.05 > * ≥ 0.01 > ** ≥ 0.001 > ***). Model results are presented in Suppl. Table 4.

Correlations between the constructs all reached significance with positive values ranging from 0.284 to 0.605, suggesting an appropriate discriminant validity between the constructs. The positive correlation structure among the constructs indicates that engineering students who have higher levels for one construct also show higher levels for another construct. The largest correlations were observed between *professional roles awareness* and *competence awareness* (cor=0.605), and between *competence awareness* and *competence confidence* (cor=0.600). *Career exploration* displayed larger correlations with the two confidence constructs than with the two awareness constructs. Moreover, *career exploration* had similar correlations for the two awareness constructs, whereas those for the confidence constructs were different, showing a larger correlation with *competence confidence*. Additionally, the correlations of *professional roles awareness* with the two confidence constructs reached different magnitudes as those for *competence awareness*. Each awareness construct had its largest association with a different confidence construct, with *professional roles awareness* showing a larger correlation with *role-fit confidence* and *competence awareness* showing a larger correlation with *competence confidence*.

3.4. Several personal variables were associated with altered identity constructs

The established measurement model is used in a MIMIC model to examine construct differences for multiple background variables, i.e. gender, phase of study, parental occupation, engineering persistence, professional role interest, and migration status. These are categorical variables that are introduced in the model accordingly by testing construct differences between the variable's categories and a reference category for each variable. Adequate model fit was obtained by including these variables in the model (CFI=0.94, TLI=0.97, RMSEA=0.04, SRMR=0.06). Additionally, factor correlations and factor loadings in the measurement model did not change substantially compared to the CFA model (Suppl. Table 5). The obtained results are summarized in Figure 5 (reference categories are listed in caption) and the results for each variable are described separately in the following sections.

1. *Gender*

No gender differences were observed for any of the five constructs, suggesting a similar professional identity for these constructs.

2. *Phase of study*

The three bachelor years demonstrated multiple significant effects for all constructs compared to the master year. Students in the first, second, and third bachelor year had reduced *career exploration* compared to master students. The sequential increase of the loadings over the three bachelor years might suggest that *career exploration* increases when progressing the programme, reaching its maximum in the master year. Similarly, *competence confidence* seemed to increase for the three bachelor years. On the other hand, *professional roles awareness* and *competence awareness* were only decreased for first and second-year students, while *role-fit confidence* was only decreased for second and third-year students. Interestingly, the significant loadings seem to increase over progressive years for all constructs, except for *professional roles awareness*, which showed a higher loading for first-year students than for second-year students. Finally, transfer students and master students showed a similar professional identity with only a small reduction in *career exploration* for transfer students.

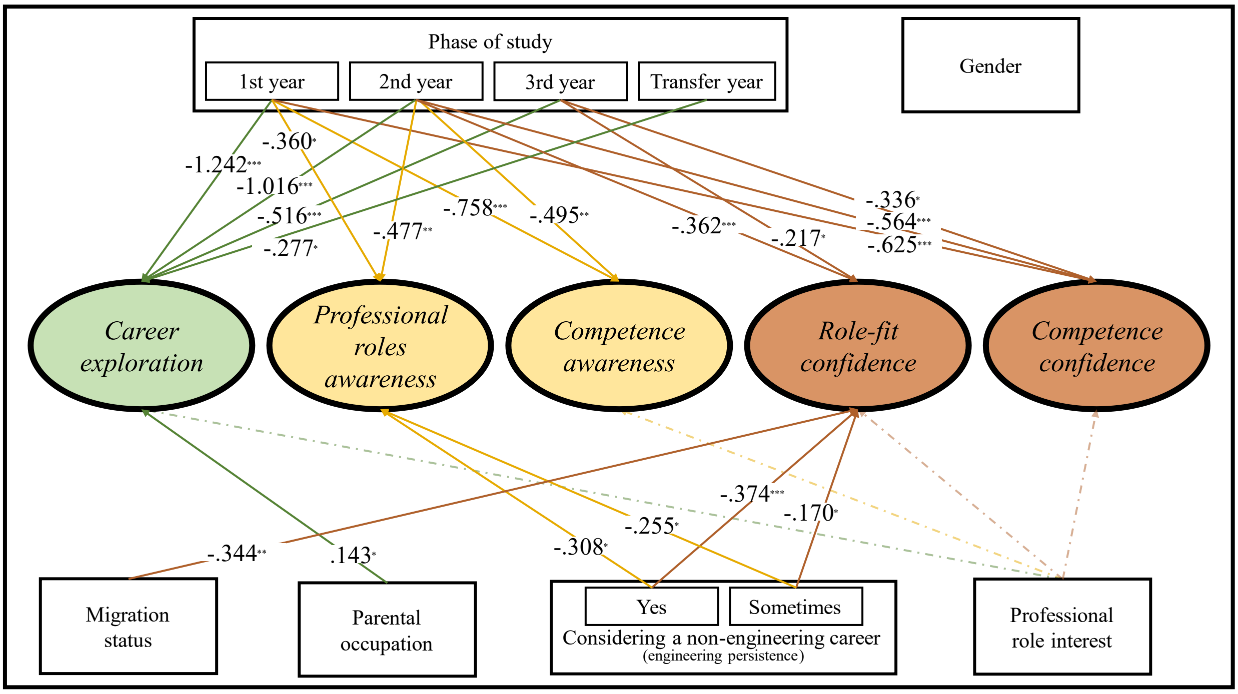


Figure 5. MIMIC model. The final CFA model was used as measurement model (Figure 4) but was here omitted for visibility. Factor loadings can be obtained in Suppl. Table 5. Results were obtained from 20 multiply imputed datasets. Reported values represent loadings based on standardized factors only and indicate the difference in the level of latent construct between the variable's category and its reference. Reference categories were the master year (phase in programme), female (gender), non-immigrant (migration status), non-engineering parent (parental occupation), and 'not considering another job than engineer' (engineering persistence). Only significant differences are displayed (0.05 > * ≥ 0.01 > ** ≥ 0.001 > ***). Dotted arrows for professional role interest indicate the presence of significant differences for the multiple comparisons for the different professional roles (Table 7).

3. Parental occupation

No substantial influence on the professional identity of engineering students was observed by whether one of their parents was an engineer or not. Only one weak effect was present that indicated slightly increased *career exploration* for students who had an engineering parent. One of the survey items for this construct involved 'talking to parents and others' (Q29), which by itself did not seem to cause the significant effect of parental occupation on *career exploration* since a significant affect was still observed when removing this item from the construct (p=0.026, loading=0.132).

4. Engineering persistence

Significant effects were only observed for two constructs, i.e. *professional roles awareness* and *role-fit confidence*. Students that considered another job outside engineering had a decreased *role-fit confidence* compared to students who aimed for an engineering career. Their *professional roles awareness* was also reduced, suggesting a lower understanding of the professional engineering roles. Additionally, students who were still undecided about their persistence in the engineering field also showed reduced levels for these two constructs.

Notably, their reductions were smaller than those for the former group (-0.255 vs. -0.308; -0.170 vs. -0.374). A borderline insignificant effect was also noted for *career exploration* (p=0.063, loading=-0.104), suggesting reduced levels in undecided students compared to persistent students.

5. Professional role interest

Construct differences were examined between eight interest groups, including seven preferred professional engineering role (PL, OE, CI, PL+OE, PL+CI, OE+CI, or PL+OE+CI) and one group indicating a current absence of interest (no idea). The construct differences between the different roles were determined by refitting the SEM with different reference groups and by applying a false discovery rate adjustment on the p-values (Table 7). Changing the reference level did not alter the results for the other variables or model fit indices (data not shown).

The multiple comparisons only showed differences in *career exploration* and *role-fit confidence* between students with different professional role preferences. Of these two, *career exploration* showed the most significant differences, which suggested that students favoring the PL or PL+OE+CI role had higher *career exploration* compared to almost all other roles. In contrast, *role-fit confidence* showed only three significant effects, of which two for PL and one for PL+CI, suggesting only a marginal influence for this construct. Additionally, students that could not yet identify with one of the engineering roles mainly showed construct differences for *role-fit confidence* and *competence confidence*. These students had a reduced *role-fit confidence* compared to all role preferences, while *competence confidence* was reduced compared to four roles. *Career exploration* and *competence awareness* were only reduced with respect to two and one professional role.

6. Migration status

No prominent differences were noted in the professional identity of students with a migration background compared to non-immigrant students. Only *role-fit confidence* was decreased in immigrant students.

As a remark, it would be interesting to examine whether a construct is measured similarly across different groups prior to investigating construct differences, for example to check if the identity constructs are represented similarly for males and females. This can be assessed by a multigroup confirmatory factor analysis for the background variables. However, such an analysis was not feasible for the current survey data because of model non-convergence. Several student groups were associated with zero counts for multiple survey item categories, which indeed impedes model convergence. Removing these zero counts by combining response categories did not resolve the convergence, suggesting a possible impact of the low group sample sizes.

Table 7. Multiple comparison of factor levels for professional role interest categories. Non-significant differences are not displayed. No significant differences were observed for *professional roles awareness*. P-values were adjusted by the false discovery rate and significance levels are indicated with asterisks (0.05 > * ≥ 0.01 > ** ≥ 0.001 > ***). The reported values represent the difference between the corresponding category (first column) and the reference level. A negative value of '-a' between group X and reference Y thus indicates that the construct level for X is decreased with a value of 'a' compared to Y, or similarly, that the level of Y is higher than that of X. Results were obtained from 20 multiply imputed datasets.

Career exploration							
	Reference level						
	PL	OE	CI	PL+OE	PL+CI	OE+CI	PL+OE+CI
PL	.						
OE	-.336**	.					-.480***
CI	-.340**		.				-.466***
PL+OE		.272*	.265*	.			
PL+CI	-.218*				.		-.344*
OE+CI	-.239*					.	-.370**
PL+OE+CI							.
No idea	-.284*						-.404**

Competence awareness							
	Reference level						
	PL	OE	CI	PL+OE	PL+CI	OE+CI	PL+OE+CI
PL	.						
OE		.					
CI			.				
PL+OE				.			
PL+CI					.		
OE+CI						.	
PL+OE+CI							.
No idea							-.874*

Role-fit confidence							
	Reference level						
	PL	OE	CI	PL+OE	PL+CI	OE+CI	PL+OE+CI
PL	.						
OE		.					
CI			.				
PL+OE	-.247*			.			
PL+CI					.		
OE+CI	-.452***				-.257*	.	
PL+OE+CI							.
No idea	-1.107***	-.829***	-.767***	-.848***	-.891***	-.635***	-.925***

Competence confidence							
	Reference level						
	PL	OE	CI	PL+OE	PL+CI	OE+CI	PL+OE+CI
PL	.						
OE		.					
CI			.				
PL+OE				.			
PL+CI					.		
OE+CI						.	
PL+OE+CI							.
No idea	-.674*	-.589*			-.666*		-.698*

3.5. Identity constructs showed several direct and mediating effects

Based on the available literature, a hypothetical path model was established to examine the magnitude of direct and indirect effects of constructs on each other. This allowed to provide a suggestive idea of the direct and mediating effects of constructs, as well as the size of the total effect of a construct on another. Model fit indices reached appropriate values upon introduction of the directional paths (CFI=0.972, TLI=0.968, RMSEA=0.046, SRMR=0.058) and all factor loadings had acceptable values above 0.30 (Suppl. Table 6).

The path model demonstrated multiple significant effects between the constructs (Figure 6A). In contrast, no direct effect was noted for *professional roles awareness* on *competence confidence*, indicating that an increased *professional roles awareness* does not affect *competence confidence*. Similarly, an insignificant path was present for *competence awareness* on *role-fit confidence*. All paths for *career exploration* were significant, suggesting that and increase in *career exploration* subsequently improves all other constructs. The two awareness constructs were affected similarly by *career exploration*, while different effects were present on the two confidence constructs with *competence confidence* increasing the most by *career exploration*. Similarly, the two awareness constructs had different effects on the two confidence constructs as a result of their insignificant effects. There, *professional roles awareness* only improved *role-fit confidence* and *competence confidence* only improved *competence confidence*.

In addition to the direct effects of *career exploration* on the confidence constructs, indirect effects were also present that were mediated by the awareness constructs (Figure 6B). First, c*areer exploration* significantly affected *role-fit confidence* via *professional roles awareness*, and *competence confidence* via *competence awareness*, indicating a sequential stimulation from exploration to awareness to confidence. In contrast, no mediation was present on *role-fit confidence* via *competence awareness*, nor on *competence confidence* via *professional roles awareness*, due to the respective insignificant awareness effects on the confidence constructs. Second, the indirect effect of *career exploration* on *competence confidence* via *competence awareness* was stronger than that on *role-fit confidence* mediated by *professional roles awareness*. Similarly, the total effect of *career exploration*, which combines the direct and mediating effects, was largest for *competence confidence* compared to *role-fit confidence*.

Finally, the direct effect of *professional roles awareness* on *role-fit confidence* has a magnitude that is comparable with that of the total effect of *career exploration* on *role-fit confidence* (0.308 vs. 0.330). This suggests a similar impact on *role-fit confidence* by direct stimulation of *professional roles awareness* as for stimulation by *career exploration*. On the other hand, the direct effect of *competence awareness* on *competence confidence* exceeded the total effect of *career exploration* on *competence confidence* (0.616 vs. 0.476).

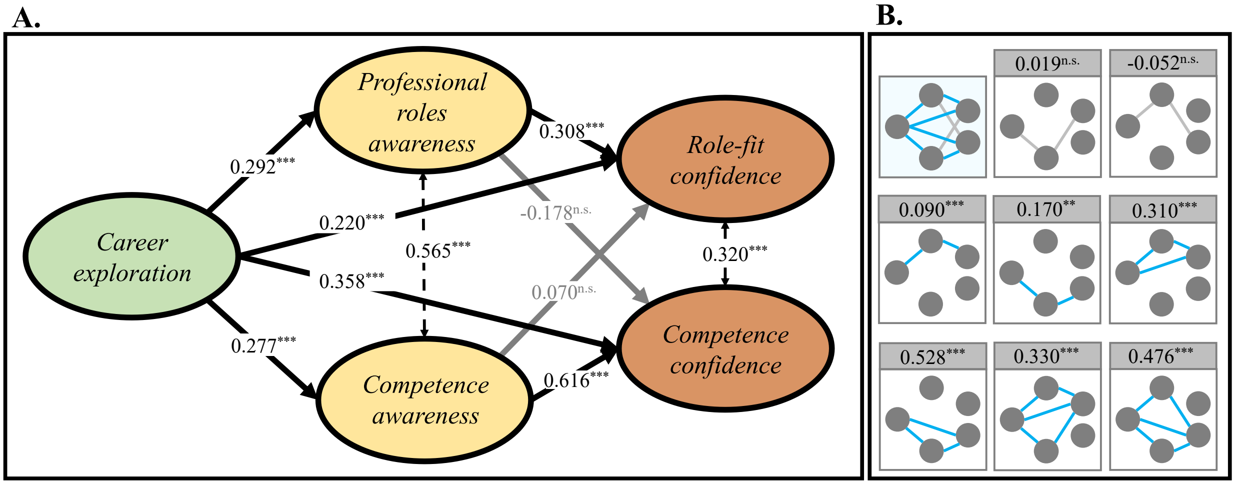


Figure 6. A Directional path model between constructs. Single-headed and double-headed arrows represent directional paths and correlations between constructs respectively. Black and grey arrows represent significant and insignificant paths respectively. **B.** Mediation path significance. The model configuration is presented schematically by the dots corresponding to the constructs in panel A. Construct connections are indicated by lines for visibility. Size and significance of aggregated paths are indicated above each panel. Color code indicates the significance of the mediation paths (blue: significant, grey: insignificant). The upper left panel summarizes significance of individual paths from panel A.

4. Discussion

Developing a professional identity is a fundamental aspect in the career orientation of students, making this an active research area underlying essential educational career guidance programmes. The present research aimed to improve the general understanding of the professional identity of engineering students, who currently lack an extensive focus as a research population despite that they have shown several career difficulties. This study thereby adds to the fundamental identity knowledge in these students, which contributes to the practical implementation of future career guidance. Using structural equation modeling techniques, we simultaneously investigated five distinct professional identity constructs in engineering students, thereby providing an in-depth exploration in the same cohort and probing an extensive identity image that has rarely been reported before. Because of the fundamental character of this work, the following discussion will focus on various research aspects, including the statistical analysis, obtained results, implications for career guidance, research limitations, and future suggestions.

4.1 Research methodology

4.1.1 Structural equation modeling

Methodological advantages

Structural equation modeling is a widely used technique in psychology (Karimi & Meyer, 2014), social science (de Carvalho & Chima, 2014), and educational science (Khine, 2013) as a confirmatory approach for a theoretical framework (Kline, 2016; Bollen, 1989). The motivation for using this method in the current research was fourfold and aimed to surpass methodological issues that are often associated with the professional identity literature. First, the use of structural equation modeling allowed the investigation of latent identity constructs that are measured indirectly by a variety of survey questions, such as the identity constructs. This improves the representation of the latent constructs as opposed to studies that employ only a single question to directly measure the respective construct (Hashish, 2019). Second, it allowed to examine five distinct identity constructs simultaneously, which adds to the current body of literature where mostly a restricted set of constructs was included at once (Patrick & Borrego, 2016). Third, this approach also recognizes that the constructs are represented by multiple measurable variables that each contribute differently to the latent construct (Bollen, 1989). Lastly, it allowed to validate the survey items by determining the representation of each item towards the constructs. Despite these final two advantages, researchers have sometimes employed less intensive methods to establish identity constructs, such as the unweighted averaging or summing of survey item scores (Cech *et al*., 2011; Creed *et al*., 2007; Jaiswal & Choudhuri, 2017; Rayman *et al*., 1983; Taveira *et al*., 1998). This improperly establishes the construct by disregarding both the unique contribution of each variable and the possible redundancy of items since the averaging follows the prior belief that the items are true representatives of the construct. The current research avoids these issues by using structural equation modeling techniques that facilitated both the retention of survey items that adequately represented the constructs, and the incorporation of item-specific construct contributions.

Causality is not inferred by the presented models

Since the latent constructs are unobservable, a good validation of the survey items is required to ensure that the hypothesized constructs are accurately represented in the analysis. It is common practice to achieve this according to various model fit indices that validate different aspects of the model (Brown, 2006, p.82-87; Kline, 2016, p.269). Crude cut-off values for each index have been proposed in literature, however, these have fluctuated according to the context of the study and remain prone to subjectivity (Svetina *et al*., 2019). Values above 0.95 for CFI and TLI, and below 0.05 or 0.06 for SRMR and RMSEA have been suggested mostly in the setting of continuous data with a reasonable number of items (Brown, 2006, p.87; Hooper *et al*., 2008). However, in the field of sociology and educational science, cut-off values are often relaxed to values of 0.90 and 0.08 (Bosscher & Smit, 1998; Brown, 2006, p.87; Hair *et al*., 2010; Laverdière *et al*., 2013). Jiang (2016) even reported an RMSEA value of 0.09 to justify a job-fit model, Doménech-Betoret *et al*. (2017) have reported a CFI of 0.897 as suggestive for the reasonable fit of their academic self-efficacy model, and Bosscher & Smit (1998) have reported an SRMR of 0.10 and CFI of 0.83. Based on values acceptable in sociology, our developed model showed good fit in all analyses, suggesting the suitability of the survey in representing the identity model.

However, the appropriate model fit does not provide evidence for the assumed underlying causal mechanisms for the effect of background variables on constructs, or the direct effects between constructs. Causality in structural equation modeling has been debated widely (Bollen, 2013; Harley *et al*., 1997), stating that “researchers do not derive causal relations from an [*sic*] SEM. Rather, the SEM represents and relies upon the causal assumptions of the researcher [which] derive from the research design, prior studies, scientific knowledge, logical arguments […] and other evidence” (Bollen, 2013, p.309). This first implied the need for supportive research for the development of a credible path model that provides insights in the effect of the constructs on each other, which had not been reported before. Supporting literature was sought with a prime interest on intervention studies that assessed construct changes after the stimulation of another construct, as suggestive for a direct effect between them. This resulted in a sensible hypothesized model that presented an intuitive directionality originating at *career exploration* and that moves downstream to awareness and confidence. However, the hypothesized model could suffer from extrapolations issues since the retrieved intervention studies were mainly performed in mixed student populations, high school students, and non-engineering students in different countries. This renders the hypothesized model explorative in nature rather than a proof for the complex interplay and mediations in the professional identity of the engineering students. Second, similarly, the effects of the background variables on the constructs in the MIMIC model should be interpreted cautiously. For gender, parental occupation, and migration status, a causal effect on the constructs could be inferred since an altered professional identity cannot influence these three variables. However, such a reverse effect might be present for the others, therefore, those observed effects should be interpreted as associations instead of causal influences.

4.1.2 The survey design and its implications

Validated and non-validated survey items

The use of standardized survey questions is widely implemented in social sciences (Pauldine *et al*., 2017; Stumpf *et al*., 1983; Westbrook & Parry-Hill, 1973), and has the benefit of reliability, research productivity, and comparability. However, this is not always possible (Patrick & Borrego, 2016; Morelock, 2017), such as in the current research where the items were designed toward the specific setting of engineering and the professional engineering roles model that was developed only recently (Craps *et al*., 2021). Survey items for two of the five constructs were closely adopted from literature (Cech *et al.*, 2011), i.e. for *role-fit confidence* and *competence confidence*, while the items for the other constructs were designed based on the general impression of the available literature. Although the final model results demonstrated a good model fit, several factor loadings and Cronbach's alpha values did not reach values as high as for validated constructs reported throughout literature. This does not impede the findings of the current research, but it does impose a restriction on the implementation of several of these designed items in future research since high reliability is expected for developed scales in high-impact research. Often, factor loadings for validated survey items exceed a magnitude of 0.50 (Coleman *et al*., 2011; Hamlet *et al*., 2021) and Cronbach's alphas reach values above 0.70 (Ponterotto & Ruckdeschel, 2007), however, lower values in sociology have been reported before that coincide with our reported values (Bosscher & Smit, 1998; Shevlin & Millar, 2006). It should also be noted that a lower number of items included in a construct reduces the Cronbach's alpha value (Ponterotto & Ruckdeschel, 2007), which might add to the low values reported here for some constructs since their mean item correlations reached substantial values (Bosscher & Smit, 1998; Clark & Watson, 1995), and most factor loadings had a value above 0.50. The smaller loadings for some items and the low Cronbach's alpha values in our research raise the desire to develop new items that represent the constructs better to meet the standards in identity literature. Future research could therefore focus on the development of validated survey questions that can be efficiently used in the setting of professional engineering identity, and which will additionally improve comparability between such studies.

Survey item Likert scales

The design of the survey items, and correspondingly the collected data characteristics, determined the subsequent analysis in this study. Items were composed of different Likert scales ranging from 2-point to 5-point scales, which has been observed before in identity literature where different scales or short scales were implemented in surveys (Currie, 1975; Doménech-Betoret *et al*., 2017; Rogers *et al*., 2008; Taveira *et al*., 1998). Although structural equation modeling allows the simultaneous incorporation of different scales (Li, 2021), the small scale sizes present some implications for the analysis, and correspondingly, for the research design. First, the survey data cannot be considered as continuous data, requiring the use of an ordinal categorical estimation technique (Muthén, 1984), with an appropriate estimator such as the recommended DWLS estimator (Bandalos, 2014; Forero *et al*., 2009; Li, 2016a & 2016b; Svetina *et al*. 2019). Future survey designs could consider implementing larger Likert scales, since larger scales are considered more efficient in analyses (Awang *et al*., 2016). Second, the categorical nature of the data is associated with more difficult model convergence, increasing the demand for larger sample sizes than for continuous data. Various sample size proposals have been discussed before, however, these mainly regarded continuous data. Nevertheless, there is a general agreement that the proposals are only suggestive and that the specific nature of the research should be considered since required sample sizes depend on various model aspects, such as model complexity, the number of items, item reliability, and the estimation technique (Kline, 2016, p.15; Wolf *et al.*, 2003). Therefore, proposals have taken into account the sample size (N) with respect to the number of model parameters (q), with suggestions ranging from 10:1 to 20:1 (N:q) (Kline, 2016, p.16). For categorical data no additional concrete suggestions have been distinguished, although larger sample sizes are generally assumed. Our reported analyses converged properly with a sample size of 624 students, providing an N:q ratio of 16:1 for the confirmatory factor analysis and the path model, and a 5:1 ratio for the structural model with the background effects. However, additional analyses that had a lower ratio failed to converge properly, such as a multigroup analysis to examine if constructs are measured similarly for males and females, probably because of low sample sizes in the respective groups. Future research could include larger Likert scales to reduce the need for large sample sizes which could facilitate additional analyses.

4.2 Results

4.2.1 Support of the survey for identity constructs

Professional identity is widely acknowledged as a broad and complex multifaceted concept that encompasses various constructs such as *career exploration*, *self-awareness*, *career awareness*, or *career confidence*. The current survey was developed with the aim to examine such a global understanding of the professional identity in engineering students by measuring these constructs with a survey. However, the exploratory results indicated that the survey does not support the putative construct of *self-awareness* as its designed survey items did not constitute a confined factor, but they either loaded with the confidence constructs or had to be removed because of low model compliance. Throughout literature, *self-awareness* is considered an important identity quality, where it has been included as a distinct construct often with a different focus, such as internal or external *self-awareness* (Özek & Ferraris, 2018), or *self-awareness* of task performance or cognitive ability (Demetriou *et al*., 2020). In our study, *self-awareness* was operationalized in the context of engineering probing their desires, abilities and how well the students could envision themselves as engineers in general. The results suggest that this construct seems indistinguishable from the confidence constructs, which probed the students' confidence that a specific engineering role is suitable for them or their confidence in the associated competencies. Multiple mechanisms could underly the inability of *self-awareness* and confidence to co-emerge. First, these constructs could indeed be indistinct in practice. This is up for future debate with psychology experts since no explicit literature was found that compares these two constructs. Also, no professional identity research was retrieved that included both *self-awareness* and *career confidence* simultaneously. Second, as suggested by the loading of several *self-awareness* questions on the confidence constructs, the design of the original questions could have corresponded too much with that of the confidence questions. Third, the designed items could intrinsically be poor representatives for the *self-awareness* construct. This might result from the use of non-validated survey questions since the current items were specifically designed in the context of engineering *self-awareness*. Future investigation could evaluate new survey items for *self-awareness* to ensure that this identity construct can be operationalized in following studies.

On the other hand, a total of five distinct constructs were supported by the survey, including subconstructs for the initial constructs of *career awareness*, i.e. *professional roles awareness* and *competence awareness*, and for *career confidence*, i.e. *role-fit confidence* and *competence confidence*. These two subconstructs for *career confidence* were reported before by Cech *et al.* (2011) and results show that they were distinguishable from the other constructs in the current survey. Additionally, various subconstructs for *career awareness* have been reported before as well, such as four different application-specific subconstructs by Dağyar *et al*. (2020). However, the specific constructs of *professional roles awareness* and *competence awareness* in engineering following the recent work of Craps *et al*. (2021) were not described before in identity research, nevertheless, the distinguishment of awareness regarding professional opportunities and the understanding of their requirements has been declared before in general *career awareness* (Watts, 2006). Despite the sensible contribution of these two retrieved subconstructs to *career awareness*, one might reconsider their practical implementation in the survey because of a strong item similarity between the questions for *competence awareness*. Namely, a negative phrasing was present for three of its four questions, which could mediate the formation of *competence awareness* as an artifact construct (Brown, 2006, p.47). This construct could indeed result from such a mechanism as further supported by the observation that an additional negatively phrased item originally loaded on this construct, which was subsequently removed due to low correspondence with the construct interpretation. Future identity research aiming to investigate *competence awareness* could redevelop non-negatively phrased questions and determine whether this construct is still distinguishable from *professional roles awareness*.

4.2.2 Professional identity in engineering students

Previous researchers have highlighted the complexity of professional identity development by investigating the interrelations of various identity constructs and by identifying several influential personal variables. Our results rehighlight this complexity on both aspects and support several implications for career guidance in the less investigated research population of engineering students. In the following sections, the construct interplay is discussed first, followed by the effects of multiple background variables.

The interplay between identity constructs

The complex construct interplay underlying professional identity is stressed by our results from the observed positive association structure among the five constructs. Previous identity research had also postulated associations between various constructs, but these show the collective disadvantage of targeting only a restricted set of identity constructs simultaneously and of focusing on different student populations. The current research investigated multiple constructs among the same students, thereby circumventing the potential extrapolation issue between cohorts with different cultural or personal characteristics (Hsieh & Huang, 2014; Özek & Ferraris, 2018). Despite this possible extrapolation issue, the results suggest that similar mechanisms might underly the identity structure of engineering students compared to other student populations since positive construct associations were also described for high school students and non-engineering students (Hsieh & Huang, 2014; Rogers *et al.*, 2008).

The positive association structure implies that increased levels for one construct are associated with higher levels for another construct. This presents a possible framework for career guidance where specific constructs could be improved by targeting others, which has already proven its efficacy in previous career development learning programmes (Hashish, 2019; Reddan, 2015). However, the reported construct associations fail to provide insights in the direction of construct stimulation, which is fundamental for developing a dedicated guidance strategy. Therefore, a directional path model was developed based on available literature that describes directional effects between these constructs, which had not been reported before for these constructs, while the relevance of such models has already been stressed for other constructs (Balkis, 2013; Diseth *et al*., 2012; Doménech-Betoret *et al*., 2017; Xu *et al*., 2014). The effects in the path model showed that the awareness and confidence constructs could be improved by *career exploration*, while restricted effects of awareness on confidence were present. This putative directionality in the path model could be of fundamental importance for the development of future engineering career guidance programmes.

Personal variables related to professional identity differences

A second layer of the identity complexity is signified by the presence of personal variables that affect its development. Again, research identifying such influential factors has been performed mainly on non-engineering students, presenting a lack of fundament insights in the engineering field (Patrick & Borrego, 2016). The current study adds to the current knowledge in engineering students by including six personal variables. Results affirm the complex view of identity as five of the variables showed associations with professional identity differences among engineering students. Following subsections discuss the obtained results for each variable in light of the current identity literature.

1. Gender

Gender was the only variable that did not affect any identity construct, while previous studies have shown that males in STEM and engineering have higher *career-fit confidence* and *competence confidence* than females (Cech *et al*., 2011; Correll, 2004; Hu *et al*., 2020; Seron *et al*., 2016; van Veelen *et al*., 2019). On the contrary, only marginal differences between engineering males and females haven been noted for their perceived soft-skill development (Naukkarinen & Bairoh, 2021). The opposing findings of the current research compared to previous studies could arise from differences between cohorts and the employed methodologies. On the other hand, our results seem to correspond with van Veelen *et al*. (2017) who showed that *career awareness* differed only marginally between high school boys and girls, however, their representation of the construct does not fully coincide with ours, impairing a sufficient comparison. Lastly, the observed gender indifference for *career exploration* replicates previous studies that repetitively demonstrated this finding in non-engineering students (Blustein, 1989; Gianakos, 1995; Hardin *et al*., 2006; Hu *et al*., 2020; Kracke, 1997; Rogers *et al*., 2008; Taveira *et al*., 1998).

2. Parental occupation

Parental influences have been considered key in professional identity development (Chin *et al*., 2019; Middleton & Loughead, 1993; Penick & Jepsen, 1992; Whiston & Keller, 2004), which is not recognized by our results. Only *career exploration* was increased when one of the parents was an engineer, while the other constructs were unaffected. This observed effect agrees with a German study that noted an increase in *career exploration* for high school students when their educational track corresponded with that of their parents (Kracke, 1997). Our result lacks an explanation of the mechanism that contributes to this increased exploration as a result of the parental occupation. Throughout literature, parental influences have been described in light of parental support (Rogers *et al*., 2008; El-Hassan & Ghalayini, 2019; Gagnon *et al*., 2019; Stringer & Kerpelman, 2010), and it could be hypothesized that our observed parental effect is mediated by the support accompanied by the mutual engineering interest. Nonetheless, the absence of substantial parental effects speculates that having an engineering parent does not affect the development of the professional identity after entering the engineering major.

3. Migration status

Immigrant students had a similar professional identity as students without a migration background, apart from their reduced *role-fit confidence* levels. Currently, the identity literature lacks the assessment of immigration effects on identity constructs, however, other minority research groups have been investigated before. Similar to our results, c*areer-fit* and *competence confidence* reductions were noted for minority students within the engineering field, despite their similar qualities as non-minority groups (Correll, 2004; Seron *et al*., 2016; van Veelen *et al*., 2019; Cech *et al*., 2011). They also tend to pursue careers where they perceive less segregation between the minority and non-minority groups (Fouad, 1995). This, in combination with the observed migration effect, signifies the relevance for enhancing the *role-fit confidence* of immigrant students who might be facing more career barriers.

4. Phase of study

Several construct differences were present for the phase of study, which agrees with the general perspective of the dynamic and evolving characteristic of professional identity (Super, 1980; van Hattum-Janssen & Endedijk, 2020). As noted before (Taveira *et al*., 1998), *career exploration* seemed to increase during the programme, reaching a maximum in the graduation year. Final-year students facing the work world are known to increase exploration to equip them for this transition (Stumpf *et al*., 1983). Additionally, developmental changes of the identity during education have also been described in light of a gained experience during the programme (Caza & Creary, 2016; Hall, 2004, p.158; Seron *et al*., 2016). This belief is supported by our results since gradual construct increases are present for *competence confidence*, *role-fit confidence* and *competence awareness* over the bachelor years. This corresponds with past findings showing that the belief of engineering students towards their career-fit and competences increased during their programme (Godwin & Lee, 2017; Pierrakos *et al*., 2016). It should be noted that a similar *role-fit confidence* was present for first and final-year students and that first-year students seemed to have a higher *professional roles awareness* than second-year students. This could result from the unrealistic expectations that first-year students enter their education with, as stated before (Bennett *et al*., 2016; Saunders-Smits *et al*., 2021). Finally, transfer students show a similar developed professional identity as master students, with the exception of a lower *career exploration* which could again correspond with a longer time to graduation. Similar to master students, they have completed a three-year programme and have obtained their bachelor's degree. The experiences associated with this educational progress might explain their similarity in identity development.

5. Professional role interest

Compared to the previous variables, not much attention has been invested for professional interest, even in non-engineering students, nor for engineering persistence. Blustein (1992) stated that students with a particular career in mind directed their exploration towards that specific career and ignored information regarding other tracks. Although our results indicated varying *career exploration* for several preferred engineering roles, these findings do not parallel those of Blustein because of our general perspective on *career exploration* compared to his career-specific measures. Additionally, our results showed no clear alterations for students without a clear role interest on *professional roles awareness*, *competence awareness* and *career exploration*. This suggests that their inability to identify with an engineering role might not result from a lack in role and competence understanding or exploration. However, the differences for *competence confidence* might present a potential mechanism where low confidence in one's competencies hinders the identification with a specific engineering role. Nevertheless, because of the non-causal nature of this association, the reverse is also possible where the inability to identify with a professional role has been the precedent for students to question their belief in their competencies to succeed in those roles.

6. Engineering persistence

A higher general professional identity has been postulated to relate to persistence in the engineering field (Eliot & Turns, 2011). Other studies were more directed towards specific identity constructs, showing that persisting students have higher *career-fit confidence* and *career awareness*, while their *competence confidence* was not altered (Cech *et al*., 2011; Zopiatis *et al*., 2016), which resembles our results. Additionally, similar proportions as for persistence among KU Leuven engineering students were reported by Lichtenstein *et al*. (2009) describing that 42% of engineering students were determined to remain in engineering, 44% were doubtful, and 14% planned on leaving engineering (vs. 37% - 47% - 16%). It is relevant to investigate the mechanism of persistence for the prevention of dropout as this occurs frequently in the engineering field (Trevelyan, 2019; van Hattum-Janssen & Endedijk, 2020). Previous research has linked dropout with decreased motivation (Beecham *et al*., 2007; Paura & Arhipova, 2014), lower correspondence between expectations and reality (Doménech-Betoret *et al*., 2017), and low correspondence of a career with one's interests (Fu *et al*., 2019). Although these factors are not represented in the current study, our results suggest that persistence in engineering might not be attributed by a general reduced professional identity, since only two constructs were reduced, i.e. *professional roles awareness* and *role-fit confidence*. In addition, the desire to leave engineering might be independent from the confidence in their non-technical competencies as no effects were noted on this construct, which presents the possibility that the motivation to consider another career might originate from a broad career interest, rather than a low perception of their non-technical engineering skills.

4.3 Implications for career guidance

General career guidance has proven to be efficient in various student populations (Hashish, 2019; Reddan, 2015; Solberg *et al*., 2002), and dedicated group-specific guidance has been proposed to respond to the needs of different students (Hsieh & Huang, 2014; Ochs & Roessler, 2004; Taveira *et al*. 1998). Our examination of the professional identity provides multiple implications regarding the career guidance of engineering students. First, general interventions aiming to stimulate desired constructs can rely on the direct and mediating effects in the developed path model. Practically, the stimulation of *career awareness* can be achieved by interactive sessions discussing the different engineering roles and their associated competencies, or by promoting active exploration. The importance of providing applied career information has been discussed before in engineering education (Bennett & Male, 2017), stressing the focus of framing the professional possibilities and required competencies in specific engineering contexts to convey their relevance rather than providing plein general information. On the other hand, students' confidence cannot simply be improved by providing knowledge about confidence, therefore, the path model shows that confidence interventions can rely on targeting *career exploration*, or *career awareness*, or both. The model results suggest that the stimulation of *role-fit confidence* could be equally efficient by both constructs, while promoting *competence confidence* seems most efficient via *competence awareness*. Evidently, an intervention combining both exploration and awareness would reach maximal efficiency in promoting confidence.

A second implication for career guidance is that the identification of several background variables for the professional identity advocates the need for personalized guidance. Students that are unable to identify with a desirable engineering role might benefit from a broad guidance programme. Such a broad intervention was already successfully implemented for students that were undecisive about their future career by simultaneously promoting their self-knowledge, exploration, and career awareness (Rayman *et al*., 1983). In contrast, a dedicated intervention targeting *professional roles awareness* might be more desirable for students who are reconsidering their persistence in the engineering field.

Similarly, immigrant students could benefit from a specific *role-fit confidence* intervention, for instance by promoting *professional roles awareness* and/or *career exploration*.

4.4 Limitations of the study and future perspectives

Although several functional models could be established and analyzed with the developed survey, a few complications were present. These included (1) the inability to represent the construct of *self-awareness*, (2) borderline fitting items for *competence awareness*, (3) the possibility that *competence awareness* is extracted as an artifact construct due to the grouping of negatively phrased items, and (4) the inability to assess whether constructs are measured similarly across different groups in a multigroup model, potentially because of low sample size. Future research could therefore focus on redesigning the identity survey to incorporate items with a higher construct representation to tackle the first three complications, and redesigning items that facilitate the performance of the desired analyses to address the fourth complication. First, improved construct representations can be achieved by using previously validated survey questions or by developing new items. Potential new items for *competence awareness* might be designed by removing the negative phrasing in the current items. For *self-awareness*, it has been noted before that this construct is more efficiently measured when the items probe the absence of *self-awareness* instead of its presence (Van Ewijk & Al-Aomar, 2016). These new items could then be validated according to proposed scale development procedures (Boateng et al. 2018; Johnson & Morgan 2016). In this process, the most promising items from a pool of items in a factor analysis could be selected and subsequently validated in the presence of other constructs to examine their discrepancy with the other constructs. Second, providing larger Likert scales might remove the need for categorical estimation techniques, thereby potentially lowering the required sample size for analyses and facilitating analyses that include a large amount of model parameters, such as multigroup analyses.

In addition to the survey design, several future research perspectives can be suggested that build upon the present findings. First, our results identified several background variables that altered the professional identity of engineering students. However, these results fail to present mechanistical evidence that constitute these identity differences, in particular for parental occupation, migration status, professional role interest and engineering persistence. Addressing this in future research that includes more directed measures, such as parental support, perceived career threat, or motivation, could greatly extend the understanding of professional identity development in these students. Second, the construct differences noted for the phases in the study programme are only suggestive for developmental changes over time as the cross-sectional data does not support a longitudinal examination. Such longitudinal identity development could be addressed in a future study that follows construct evolutions in the same individuals over time. Third, the developed path model between the constructs was based on various identity literature and could suffer from extrapolation problems, making this model only explorative. Additionally, bidirectionality between other identity constructs has been postulated before (El-Hassan & Ghalayini, 2019; Guan *et al*., 2017; Pinxten *et al*., 2009), but is not represented in the current path model, which limits a presumable complete representation of the interplay constituting the identity. Ideally, the development of a reliable path model for engineering students should be based on multiple intervention studies that assess construct alterations in these students after the stimulation of another construct. However, the required time and effort to achieve this makes this unfeasible. Fourth, the developed path model provides important practical insights for career guidance, however, the suggested construct effects are only presented in a general context. Personalized career guidance would benefit from further examination of path differences between student groups. Finally, future research should focus on developing career guidance programmes directed to engineering students and evaluate their efficacy in improving the included five identity constructs.

5. Conclusion

This study extended the general view of the professional identity of engineering students using structural equation modeling on survey data. We validated the designed survey in measuring five latent identity constructs and found that one construct was not supported by the survey, i.e. *self-awareness*, whereas five other constructs were retrieved, i.e. *career exploration*, *professional roles awareness*, *competence awareness*, *role-fit confidence*, and *competence confidence*. These constructs showed a considerable association structure among each other, which is suggestive for their complex interplay. Direct effects between constructs were examined using a literature based hypothesized path model, further examining the complex interactions among constructs and providing practical insights for the development of career interventions. Additionally, several personal variables were identified that showed an altered professional identity. Of these, the phase in the study programme and a lack of professional role interest showed the most effects on the identity. Other variables were engineering persistence, parental occupation, and migration status, which showed only a few effects, while gender did not seem to affect the identity. These results contribute to the general professional identity knowledge of engineering students, which currently lack a comprehensive understanding. The insights obtained from this work support implications for the design and implementation of future career guidance in engineering education to tackle the career orientation difficulties observed among engineering students. Finally, future studies could mainly focus on improving the survey design and exploring the mechanisms underlying the developed professional identity.

References

Awang Z, Afthanorhan A, Mamat M. The Likert scale analysis using parametric based structural equation modeling (SEM). Computation Methods in Social Sciences (2016). 13-21.

Balkis M. Academic procrastination, academic life satisfaction and academic achievement: the mediation role of rational beliefs about studying. Journal of Cognitive and Behavioral Psychotherapies (2013). 13(1):57-74.

Bandalos DL. Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. Structural Equation Modeling: A Multidisciplinary Journal (2014). 21(1):102-116.

Beecham S, Baddoo N, Hall T, Robinson H, Sharp H. Motivation in software engineering: a systematic literature review. Information and Software Technology (2008). 50:860-878.

Bennett C, Ha MR, Bennett J, Czekanski A. Awareness of self and the engineering field: student motivation, assessment of 'fit' and preparedness for engineering education. Canadian Engineering Education Association (CEEA16) Conf (2016).

Bennett D, Male SA. An Australian study of possible selves perceived by undergraduate engineering students. European Journal of Engineering Education (2017). 42(6):603-617.

Betz NE. Self-efficacy theory as a basis for career assessment. Journal of Career Assessment (2000). 8(3):205-222.

Blustein DL. The role of goal instability and career self-efficacy in the career exploration process. Journal of Vocational Behavior (1989). 35:194-203.

Blustein DL. Applying current theory and research in career exploration to practice. The Career Development Quarterly (1992). 41:174-183.

Blustein, DL., Juntunen CL, Worthington RL. The school-to-work transition: adjustment challenges of the forgotten half. In S. D. Brown & R. W. Lent (2000). (Eds.), Handbook of Counseling Psychology (3rd ed.). New York: Wiley.

Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Frontiers in Public Health (2018). 6:149.

Bollen KA. Structural equations with latent variables. John Wiley & Sons, New York (1989).

Bosscher RJ, Smit JH. Confirmatory factor analysis of the general self-efficacy scale. Behaviour Research and Therapy (1998). 36:339-343.

Bretz RD, Judge TA. Person-organization fit and the theory of work adjustment: implications for satisfaction, tenure, and career success. Journal of Vocational Behavior (1994). 44:32-54.

Brown TA. Confirmatory factor analysis for applied research. Kenny DA (Eds). The Guilford Press, New York (2006).

Carson AD, Mowsesian R. Moderators of the prediction of job satisfaction from congruence: a test of Holland's theory. Journal of Career Assessment (1993). 1(2):130-144.

Caza BB, Creary SJ. The construction of professional identity. Wilkinson A, Hislop D, Coupland C. (Eds.). Perspectives on contemporary professional work: Challenges and experiences (2016), Edward Elgar Publishing, Cheltenham UK, pp.259-285.

Cech E, Rubineau B, Silbey S, Seron C. Professional role confidence and gendered persistence in engineering. American Sociological Review (2011). 76(5):641-666.

Cherry N. Do careers officers give good advice?. British Journal of Guidance and Counselling (1974). 2(1):27-40.

Chin MY, Cohen CAB, Hora MT. Examining US business undergraduates' use of career information sources during career exploration. Education + Training (2020). 62(1):15-30.

Choe NH, Borrego M. Prediction of engineering identity in engineering graduate students. IEEE Transactions on Education (2019). 62(3):181-187.

Clark LA, Watson D. Constructing validity: basic issues in objective scale development. Psychological Assessment (1995). 7(3):309-319.

Coleman D, de Chernatony L, Christodoulides G. B2B service brand identity: scale development and validation. Industrial Marketing Management (2011). 40:1063-1071.

Correll SJ. Constraints into preferences: gender, status, and emerging career aspirations. American Sociological Review (2004). 69:93-113.

Craps S, Pinxten M, Knipprath H, Langie G. Different roles, different demands. A competence-based professional roles model for early career engineers, validated in industry and higher eduction, European Journal of Engineering Education (2021).

Creed PA, Patton W, Prideaux LA. Predicting change over time in career planning and career exploration for high school students. Journal of Adolescence (2007). 30:377-392.

Currie LE. An index of vocational awareness. Vocational Guidance Quarterly (1975). 23(4):347-353.

Dağyar M, Kasalak G, Uğurlu N. Academic career awareness and academic career interest among Turkish undergraduate students. Cypriot Journal of Educational Sciences (2020). 15(5):937-954.

de Carvalho J, Chima FO. Applications of structural equation modeling in social sciences research. American International Journal of Contemporary Research (2014). 4(1):6-11.

Demetriou A, Kazi S, Makris N, Spanoudis G. Cognitive ability, cognitive self-awareness, and school performance: from childhood to adolescence. Intelligence (2020). 79:101432.

Dik BJ, Steger MF, Gibson A, Peisner W. Make your work matter: development and pilot evaluation of a purpose-centered career education intervention. New Directions for Your Development (2011). 132:59-73.

Diseth A, Danielsen AG, Samdal O. A path analysis of basic need support, self-efficacy, achievement goals, life satisfaction and academic achievement level among secondary school students. Educational Psychology (2012). 32(3):335-354.

Doménech-Betoret F, Abellán-Roselló L, Gómez-Artiga A. Self-efficacy, satisfaction, and academic achievement: the mediator role of students' expectancy-value beliefs. Frontiers in Psychology (2017). 8:1193.

El-Hassan K, Ghalayini N. Parental attachment bonds, dysfunctional career thoughts and career exploration as predictors of career decision-making self-efficacy of Grade 11 students. British Journal of Guidance & Counselling (2019).

Eliot M, Turns J. Constructing professional portfolios: sense-making and professional identity development for engineering undergraduates. Journal of Engineering Education (2011). 100(4):630-654.

Flaherty EA, Day CC, Urbanek RE, Wood DM, D'acunto LE, Quinn VS, Zollner PA. Mentored conference experience support student's career exploration and professional development. Wildlife Society Bulletin (2019). 43(3):565-575.

Forero CG, Maydeu-Olivares A, Gallardo-Pujol D. Factor analysis with ordinal indicators: a monte carlo study comparing DWLS and ULS estimation. Structural equation modeling (2009). 16:625-641.

Fouad NA. Career linking: an intervention to promote math and science career awareness. Journal of Counseling & Development (1995). 73:527-534.

Fu M, Zhang LF, Li B. Revisiting the congruence-satisfaction relationship: the role of external forces. Journal of Career Development (2019). 46(3).203-218.

Gagnon É, Ratelle CF, Guay F, Duchesne S. Developmental trajectories of vocational exploration from adolescence to early adulthood: The role of parental need supporting behaviors. Journal of Vocational Behavior (2019). 115:103338.

Gati I, Levin N. Counseling for career decision-making difficulties: measures and methods. The Career Development Quarterly (2014). 62:98-113.

Gianakos I. The relation of sex role identity to career decision-making self-efficacy. Journal of Vocational Behavior (1995). 46:131-143.

Godwin A, Lee WC. A cross-sectional study of engineering identity during undergraduate education. School of Engineering Education Faculty Publications (2017).

Gu X, Tang M, Chen S, Montgomery MLT. Effects of a career course on Chinese high school student's career decision-making readiness. The Career Development Quarterly (2020). 68:222-237.

Guan Y, Zhuang M, Cai Z, Ding Y, Wang Y, Huang Z, Lai X. Modeling dynamics in career construction: reciprocal relationship between future work self and career exploration. Journal of Vocational Behavior (2017). 101:21-31.

Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis, 7th ed (2010), Pearson Prentice Hall, NJ.

Hall DT. Self-awareness, identity, and leader development. Day DV, Zaccaro SJ, Halpin SM (Eds.). Leader development for transforming organizations. Lawrence Erlbaum Associates, New Jersey (2004).

Hamlet LC, Roy A, Scalone G, Lee R, Poleacovschi C, Kaminsky J. Gender and engineering identity among upper-division undergraduate students. Journal of Management in Engineering (2021). 37(2):04020113.

Hardin EE, Varghese FP, Tran UV, Carlson AZ. Anxiety and career exploration: gender differences in the role of self-construal. Journal of Vocational Behavior (2006). 346-358.

Harley AE, Scandura TA, Schriesheim CA, Brannick MT, Seers A, Vandenberg RJ, Williams LJ. Exploratory and confirmatory factor analysis: guidelines, issues, and alternatives. Journal of Organizational Behavior (1997). 18:667-683.

Hashish EAA. The effect of career awareness on perceived career and talent development self-efficacy and career barriers among nursing students. Journal of Research in Nursing (2019). 24(3-4):233-247.

Hofland E, Pinxten M, Wauters D, Langie G (2015). Roles in the bachelor's and master's programmes in engineering technology. Proceedings of the 43rd SEFI annual conference 2015 – Diversity in engineering education: an opportunity to face the new trends of engineering.

Hooper D, Coughlan J, Mullen MR. Structural equation modelling: guidelines for determining model fit. The Electronic Journal of Business Research Methods (2008). 6(1):53-60.

Horn JL. A rationale and test for the number of factors in factor analysis. Psychometrica (1965). 30(2):179-185.

Hsieh HH, Huang JT. The effects of socioeconomic status and proactive personality on career decision self-efficacy. The Career Development Quarterly (2014). 62:29-43.

Hu S, Hood M, Creed PA, Shen X. The relationship between family socioeconomic status and career outcomes: a life history perspective. Journal of Career Development (2020). 1-16.

Inkson K, Elkin G. Landscape with travellers: the context of careers in developed nations. Athanasou JA, Van Esbroeck R (Eds.). International handbook of career guidance (2008). Springer,.

Jaiswal SK, Choudhuri R. Academic self concept and academic achievement of secondary school students. American Journal of Educational Research (2017). 5(10):1108-1113.

Jiang Z. The relationship between career adaptability and job content plateau: the mediating roles of fit perceptions. Journal of Vocational Behavior (2016). 95-96:1-10.

Johnson RL, Morgan GB. Survey scales: a guide to development, analysis, and reporting. The Guilford Press, New York (2016).

Jorgensen TD, Pornprasertmanit S, Schoemann AM, Rosseel Y, (2018). SemTools: Useful tools for structural equation modeling. R package version 0.5-4. Retrieved from https://CRAN.R-project.org/package=semTools

Kaiser HF, Rice J. Little Jiffy, Mark IV. Educational and Psychological Measurement (1974). 34:111-117.

Karbownik MS, Horne R, Paul E, Kowalczyk E, Szemraj J. Determinants of knowledge about dietary supplements among polish internet users: nationwide cross-sectional study. Journal of Medical Internet Research (2021). 23(4):e25228.

Karimi L, Meyer D. Structural equation modeling in psychology: the history, development and current challenges. International Journal of Psychological Studies (2014). 6(4):123-133.

Keumala E, Nurihsan J, Budiamin A. The development of career learning program with modeling techniques to improve student career awareness. Islamic Guidance and Counseling Journal (2018). 1(2):53-61.

Khine MS. Application of structural equation modeling in educational research and practice. Sense Publishers, Rotterdam (2013).

Kline RB. Principles and practice of structural equation modeling. Kenny DA, Little TD (Eds). The Guilford Press, New York, 4th edition (2016).

Kosby SI, Mariano FM. Promoting youth purpose: a review of the literature. New Directions for Youth Development (2011). 132:13-29.

Kracke B. Parental behaviors and adolescent's career exploration. The Career Development Quarterly (1997). 45:341-350.

Kracke B. The role of personality, parents and peers in adolescents career exploration. Journal of Adolescence (2002). 25:19-30.

Langie G, Craps S. Professional competencies in engineering education: the PREFERed-way. Információs Társadalom XX (2020), no. 2:142-153.

Laverdière O, Morin AJS, St-Hilaire F. Factor structure and measurement invariance of a short measure of the big five personality traits. Personality and Individual Differences (2013). 55:739-743.

Lent RW, Morris TR, Penn LT, Ireland GW. Social-cognitive predictors of career exploration and decision-making longitudinal test of the career self-management model. Journal of Counseling Psychology (2019). 66(2):184-194.

Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. Behavior Research Methods (2016a). 48:936-949.

Li CH. The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. Psychological Methods (2016b). 21(3):369-387.

Li CH. Statistical estimation of structural equation models with mixture of continuous and categorical observed variables. Behavior Research Methods (2021).

Lichtenstein G, Loshbaugh HG, Claar B, Chen HL, Jackson K, Sheppard SD. An engineering major does not (necessarily) an engineer make: career decision making among undergraduate engineering majors. Journal of Engineering Education (2009). 98(3):227-234.

Lie D, Trial J, Schaff P, Wallace R, Elliott D. "Being the best we can be": medical student's reflections on physician responsibility in the social media era. Academic Medicine (2013). 88(2):240-245.

Lorenzo-Seva U, Ferrando PJ. Non positive definite correlation matrices in exploratory item factor analysis: causes, consequences and a proposed solution. Structural equation modeling: a multidisciplinary Journal (2020).

Matusovich H, Streveler R, Miller R, Olds B. I'm graduating this year! So what is an engineer anyway? Americal Society of Engineering Education (2009).

Middleton EB, Loughead TA. Parental influence on career development: an integrative framework for adolescent career counseling. Journal of Career Development (1993). 19(3):161-173.

Morelock JR. A systematic literature review of engineering identity: definitions, factors, and interventions affecting development, and means of measurement. European Journal of Engineering Education (2017). 42(6):1240-1262.

Muthén B. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika (1984). 49(1):115-132.

Naukkarinen J, Bairoh S. Gender differences in professional identities and development of engineering skills among early career engineers in Finland. European Journal of Engineering Education (2021).

Nauta MM. Career interests, self-efficacy, and personality as antecedents of career exploration. Journal of Career Assessment (2007). 15(2):162-180.

Ochs LA, Roessler RT. Predictors of career exploration intentions: a social cognitive career theory perspective. Rehabilitation Counseling Bulletin (2004). 47(4):224-233.

Özek H, Ferraris A. The role of self-awareness and self-efficacy on career decision making: an innovative perspective. Vrontis D, Weber Y, Thrassou A. (Eds). Innovation and capacity building (2018), Palgrave Macmillan, Cham Switzerland, pp. 207-225.

Patrick AD, Borrego M. A review of the literature relevant to engineering identity. ASEE's 123rd Annual (2016), Conference & Exposition, New Orleans, LA.

Pauldine MR, Warren JA, Swails JA. Tennessee self-concept scale. Ziegler-Hill V, Shackelford T (Eds). Encyclopedia of Personality and Individual Differences (2017). Springer, Cham.

Paura L, Arhipova I. Cause analysis of student's dropout rate in higher education study program. Procedia Social and Behavioral Sciences (2014). 109:1282-1286.

Penick NI, Jepsen DA. Family functioning and adolescent career development. The Career Development Quarterly (1992). 40:208-222.

Pierrakos O, Curtis NA, Anderson RD. How salient is the identity of engineering students? On the use of engineering student identity survey. IEEE Frontiers in Educational Conference (FIE) (2016).

Pinxten M, De Fraine B, Van Damme J, D'Haenens E. Development of the relation between achievement and academic self-concept in secondary school. Annual Meeting of the American Educational Research Association, San Diego, CA (2009).

Ponterotto JG, Ruckdeschel DE. An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. Perceptual and Motor Skills (2007). 105:997-1014.

R Core Team R: A language and environment for statistical computing. (2019). R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raykov T. Behavioral scale reliability and measurement invariance evaluation using latent variable modeling. Behavior Therapy (2004). 35:299-331.

Rayman JR, Bernard CB, Holland JL, Barnett DC. The effect of a career course on undecided college students. Journal of Vocational Behavior (1983). 23:346-355.

Reddan G. Enhancing student's self-efficacy in making positive career decisions. Asia-Pacific Journal of Cooperative Education (2015). 16(4):291-300.

Revelle W. Package 'psych' (2014). Retrieved from http://cran.r-project.org/web/packages/psych/psych.pdf

Rogers ME, Creed PA, Glendon AI. The role of personality in adolescent career planning and exploration: a social cognitive perspective. Journal of Vocational Behavior (2008). 73:132-142.

Saunders-Smits G, Craps S, Carthy D, Langie G. Comparison of first-year student conceptions of their future roles as engineers between Belgium, Ireland, and The Netherlands. International Journal of Mechanical Engineering Education (2021). 0(0):1-19.

Savickas ML, Nota L, Rossier J, Dauwalder JP, Duarte ME, Guichard J, Soresi S, Van Esbroeck R, van Vianen AEM. Life designing: a paradigm for career construction in the 21st century. Journal of vocational behaviour (2009). 75:239-250.

Schmitt-Rodermund E. Breadth of interests, exploration, and identity development in adolescence. Journal of Vocational Behaviour (1999). 55:298-317.

Schneider B, Stevenson D. The Ambitious Generation: America's Teenagers, Motivated but Directionless. New Haven, CT: Yale University Press (1999).

Seron C, Silbey SS, Cech E, Rubineau B. Persistence is cultural: professional socialization and the reproduction of sex segregation. Work and Occupations (2016). 43(2):178-214.

Shevlin M, Millar R. Career education: an application of latent growth curve modelling to career information-seeking behavior of school pupils. British Journal of Educational Psychology (2006). 73:141-153.

Skorikov VB, Vondracek F. Occupational identity. Schwartz SJ, Luyckx K, Vignoles VL (Eds). Handbook of identity theory and research (693-714). New York. Springer (2011).

Solberg VS, Howard KA, Blustein DL, Close W. Career development in the schools: connecting school-to-work-to-life. The Counseling Psychologist (2002). 30:705-725.

Stringer KJ, Kerpelman JL. Career identity development in college students: decision making, parental support, and work experience. Identity: An International Journal of Theory and Research (2010). 10(3):181-200.

Stumpf SA, Colarelli SM, Hartman K. Development of the career exploration survey (CES). Journal of Vocational Behavior (1983). 22:191-226.

Super DE. A life-span, life-space approach to career development. Journal of Vocational Behaviour (1980). 16:282-298.

Sutono SB. Selection of representative Kansei objectives using cluster analysis: a case study on car design. International Journal of Advanced Engineering, Management and Science (2016). 2(11):1885-1891.

Svetina D, Rutkowski L, Rutkowski D. Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using Mplus and the lavaan/semTools packages. Structural Equation Modeling: A Multidisciplinary Journal (2019). 0:1-20.

Taveira MDC, Silva MC, Rodríguez ML, Maia J. Individual characteristics and career exploration in adolescence. British Journal of Guidance & Counselling (1998). 26(1):89-104.

Timmerman ME, Lorenzo-Seva U. Dimensionality assessment of ordered polytomous items with parallel analysis. Psychological Methods (2011). 16(2):209-220.

Trevelyan J. Transitioning to engineering practice. European Journal of Engineering Education (2019). 44(6):821-837.

van Buuren S, Groothuis-Oudshoorn K, (2011). Mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software, Vol 45, pp. 1-67.

Van Eeghen C, Hitt J, King JG, Okech JEA, Rouleau B, Melekis K, Kessler R, Pinckney R. An interprofessional education pilot program on screening, brief, intervention, and referral to treatment (SBIRT) improves student knowledge, skills, and attitudes. International Journal of Higher Education (2019). 8(1):119-132.

Van Ewijk A, Al-Aomar R. Inspiration, self-awareness and entrepreneurial intentions: a mixed-method case study of postgraduate engineering students in the UAE. Journal of Entrepreneurship Education (2016). 19(1):103-126.

van Hattum-Janssen N, Endedijk MD. Professional identity development and career choices in engineering education: the added value of life history research. In J. van der Veen, N. van Hattum-Janssen, H-M. Järvinen, T. de Laet, & I. ten Dam (Eds.), Engaging, Engineering, Education: Book of Abstracts, SEFI 48th Annual Conference University of Twente (online), 20-24 September, 2020 (pp. 1297-1304). University of Twente.

van Veelen R, Sleegers PJC, Endedijk MD. Professional learning among school leaders in secondary education: the impact of personal and work context factors. Educational Administration Quarterly (2017). 53(3):365-408.

van Veelen R, Derks B, Endedijk MD. Double trouble: how being outnumbered and negatively stereotyped threatens career outcomes of women in STEM. Frontiers in Psychology (2019). 10:150.

Verbeke G, Molenbergs G, Kenward MG. A likelihood based perspective. Molenbergs G, Fitzmaurice G, Kenward MG, Tsiatis A, Verbeke G (Eds.), in Handbook of missing data methodology (2015), CRC Press, Boca Raton, US.

Vondracek FW. The construct of identity and its use in career theory and research. The Career Development Quarterly (1992). 41:130-144.

Watts AG. Career education in higher education: principles and practice. British Journal of Guidance & Counselling (1997). 5(2):167-184.

Watts AG. Career development learning and employability. The Higher Education Academy (2006).

Westbrook BW, Parry-Hill JW. The measurement of cognitive vocational maturity. Journal of Vocational Behavior (1973). 3:239-252.

Whiston SC, Keller BK. The influence of the family of origin on career development: a review and analysis. The Counseling Psychologist (2004). 32(4):493-568.

Wise R, Charner I, Randour ML. A conceptual framework for career awareness in career decision-making. The Counseling Psychologist (1976). 6(3):47-53.

Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. Educational and Psychological Measurement (2013). 76(6):913-934.

Woo H, Lu J, Bang N. Professional identity scale in counseling (PISC): revision of factor structure and psychometrics. Journal of Counselor Leadership and Advocacy (2018). 5(2):137-152.

Xu H, Hou ZJ, Tracey TJG. Relation of environmental and self-career exploration with career decision making difficulties in Chinese students. Journal of Career Assessment (2014). 22(4):654-665.

Zopiatis A, Theocharous AL, Constanti P. Adult vocational decision, career satisfaction and future intention: insights from the hospitality industry. International Journal of Contemporary Hospitality Management (2016). 28(12):2696-2720.

Appendix

1. *Employed literature supporting the hypothesized path model*

H1: *Self-awareness* → *Role-fit confidence*

- Gu *et al*. (2020): Provided a career intervention in Chinese high school students that focused on the understanding of their "self-portrait" regarding their interests, personality, and abilities. Although not explicitly stated, this is representative for *self-awareness*. The study noted that this intervention stimulated "career-decision self-efficacy", which was defined as the confidence in their career vision by 'goal selection' (which could serve as a surrogate connection between *self-awareness* and our *professional role confidence*), as well as the confidence in their skills, such as 'planning' and 'problem solving' (which could be suggestive for our *competence confidence*).
- Hu *et al*. (2020) + Xu *et al*. (2014): The stimulation of *role-fit confidence* by *self-awareness* is also hypothesized from the combination of the aforementioned two papers. First, Hu *et al*. (2020) provided a direct link in their SEM from 'career exploration' to 'perceived person-job fit' (suggestive for our *role-fit confidence*). Their career exploration both included occupational exploration (such as our *career exploration*) and self-exploration, showing that self-exploration might improve *role-fit confidence*. Second, Xu *et al*. (2014) states a direct link in their SEM from 'self-exploration' to 'lack of information', which includes the lack of information about the self (suggestive for *self-awareness*). Combining these two, Hu *et al*. (2020) showed that self-exploration affects *role-fit confidence*, and Xu *et al*. (2014) showed that this might be mediated by *self-awareness*. Therefore, an effect between *self-awareness* and *role-fit confidence* is hypothesized.
- Reddan (2015) showed that an intervention targeting self-awareness, as measured by the awareness of personal strengths and weaknesses, and opportunity awareness, as measured by the self-evaluated knowledge regarding a specific occupation, increased students' career-decision self-efficacy regarding goal selection (suggestive for *role-fit confidence*), planning and problem solving (suggestive for *competence confidence*).

H2: *Career-exploration* → *Career awareness*

- Shevlin *et al*. (2006) states that "most career education programmes embody objectives that endorse the acquisition of information related to both the self and career options. This is based on the assumption that information constitutes the basis for the development of self-awareness and opportunity awareness"
- Fouad (1995) designed a science-based career intervention for high school students that provided career exploration as defined by our survey: participating in extracurricular activities, going to talks, visits, experience, which improves their 'occupational knowledge' (suggestive for *career awareness*).

- Xu *et al*. (2014) provides a direct link in their SEM from 'environmental exploration' to 'lack of information about occupations and the process' (suggestive for *career awareness*).

H3: *Career exploration* → *Self-awareness*

- Shevlin *et al*. (2006): see H2
- Xu *et al*. (2014) provides a direct link in their SEM from 'environmental exploration' to 'lack of information about the self' (suggestive for *self-awareness*).

H4: *Career awareness* → *Competence confidence*

- Hashish (2019) provided nursing students with a 'career awareness course', which improved their 'career and talent development self-efficacy' as measured by "their belief about one's own competencies and ability to succeed" (suggestive for *competence confidence*).
- Reddan (2015): see H1

H5: *Career exploration* → *Role-fit confidence*

- Hu *et al*. (2020) provided a direct link in their SEM from 'career exploration' to 'perceived person-job fit' (suggestive for our *role-fit confidence*).
- Flaherty *et al*. (2019) showed that an intervention where students had to go to conferences (which is reflected by our *career exploration*) improved their 'professional and social abilities' (suggestive for *competence confidence*) and they developed a new sense of belonging to their field (suggestive for *role-fit confidence*).
- Downing *et al*. (2010) showed a direct path from *career exploration* to 'career indecision', defined as "the inability to specify a career goal" (which might be suggestive for *role-fit confidence*).

H6: *Career exploration* → *Competence confidence*

- Flaherty *et al*. (2019): see H5
- Lau *et al*. (2019) provided a self-awareness, defined by interests an aptitudes, and career exploration based intervention that promoted 'career maturity', which partly included the attitude towards one's competences (suggestive for our *competence confidence*).

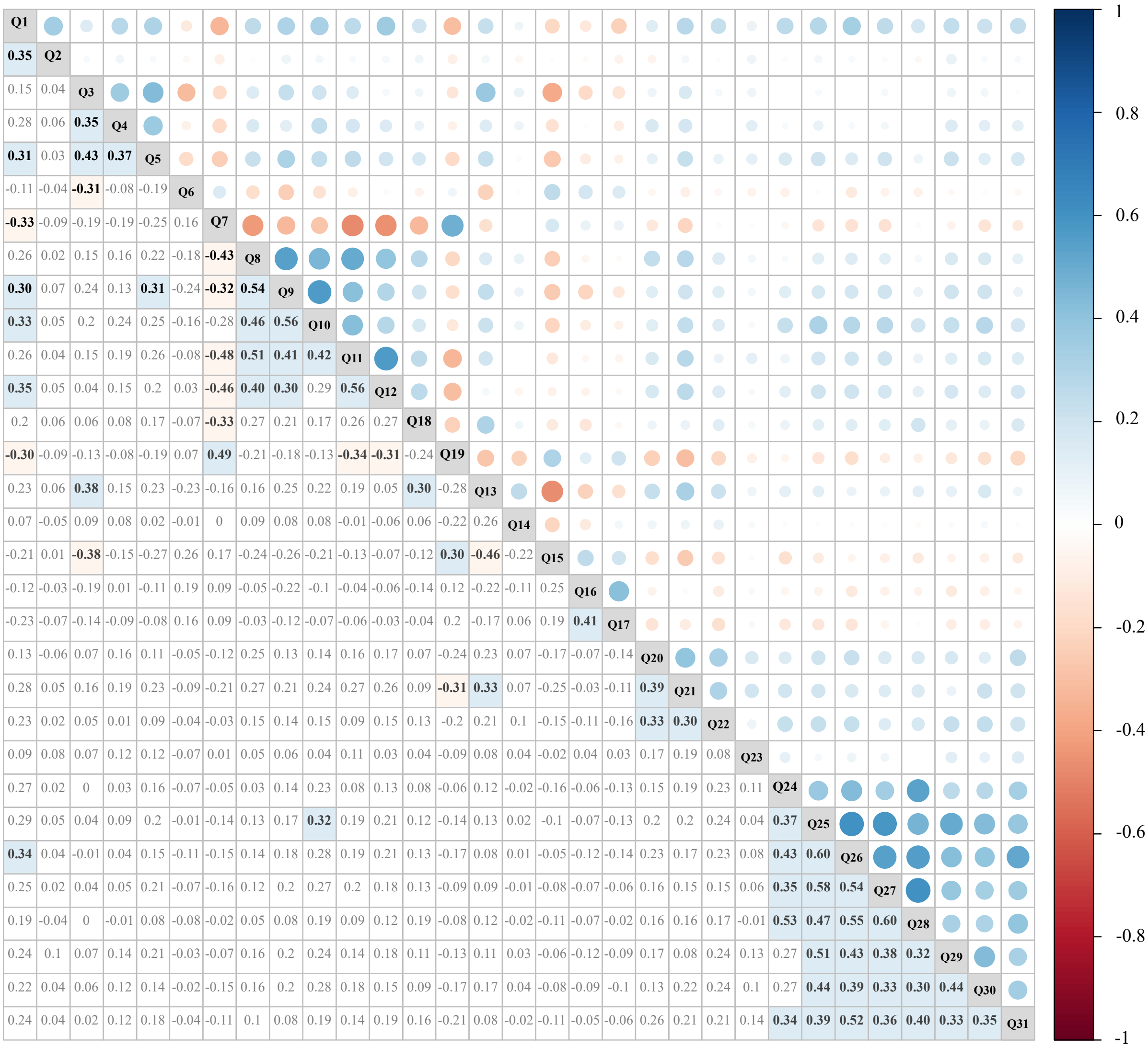
H7: *Self-awareness* → *Competence confidence*

- Lau *et al*. (2020) provides a direct effect in their SEM from 'self-concept' (suggestive for *self-awareness*) to 'work readiness', which was defined as being "not concerned regarding responsibilities, flexibility, skills and communication" (suggestive for *competence confidence*).
- Reddan (2015): see H1
- Gu *et al*. (2020): see H1

H8: *Career awareness* → *Role-fit confidence*

- Jyoti *et al*. (2015) suggested that career awareness programmes increase confidence in the job.
- Reddan (2015): see H1

2. *Supplementary figures*



Suppl. Figure 1. Polychoric correlation plot. Upper diagonal: absolute circular representation of correlation magnitudes. Lower diagonal: correlation values, absolute magnitudes exceeding 0.30 are indicated in bold and the panel color shows positive and negative values in blue and red.

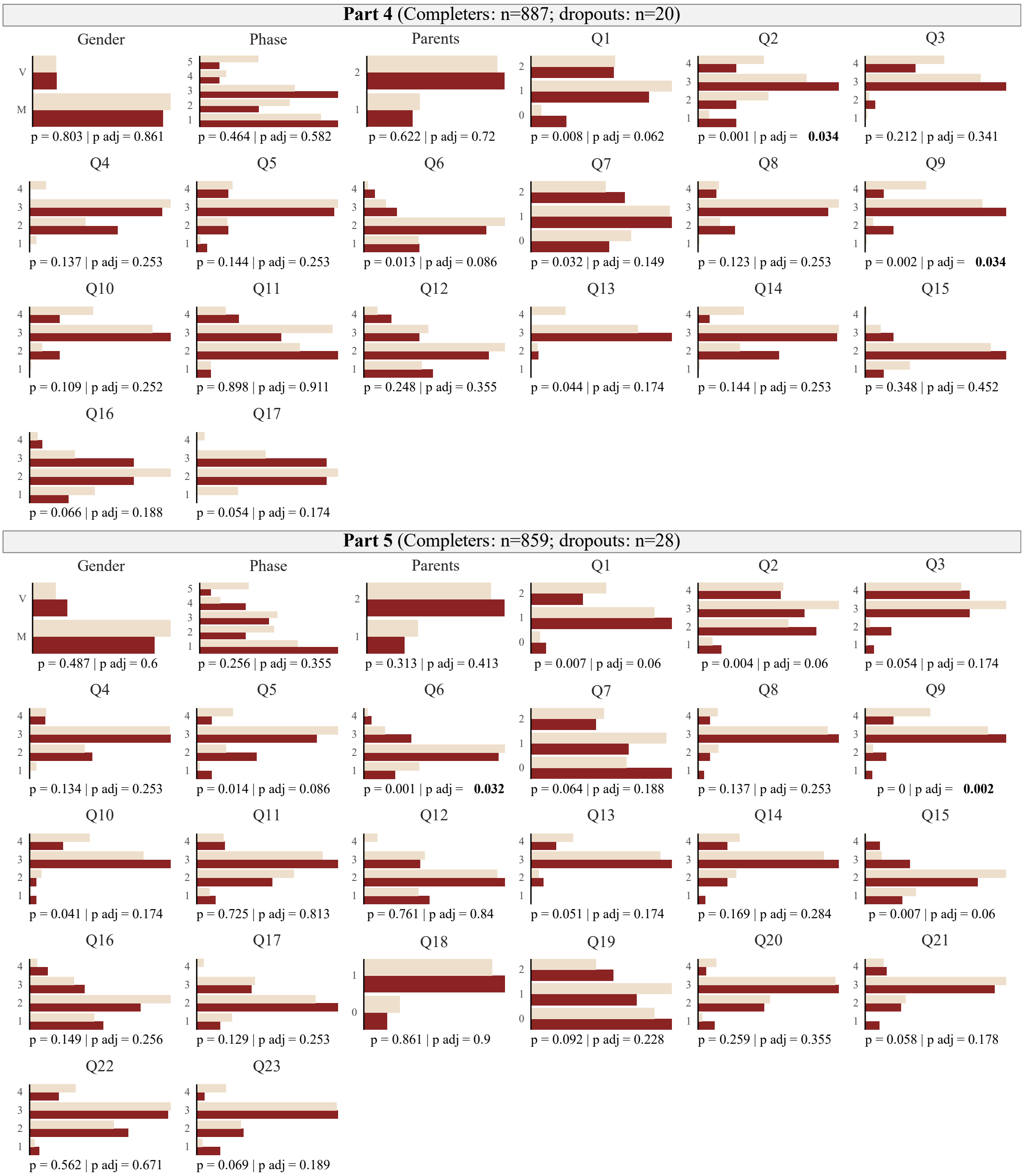
A.

	Included survey items							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Part 1	Q2	Q3	Q4	Q5	Q6			
Part 2	Q7	Q8	Q9	Q10	Q11	Q12		
Part 3	Q13	Q14	Q15	Q16	Q17			
Part 4	Q18	Q19	Q20	Q21	Q22	Q23		
Part 5	Q24	Q25	Q26	Q27	Q28	Q29	Q30	Q31

B.

(continue)

B. (cont'd)



Suppl. Figure 2. Relationship between dropout and data. A. Survey items included in each partition. **B.** Relative incidence of the variables' categories for the completers (beige) and dropouts (red) group. Therefore, all beige bars add up to 100 %, similarly for the red bars. Fisher exact tests ('p') were used for the comparisons. Adjusted p-values ('p.adj') were calculated based on the false discovery rate. Significant adjusted p-values are indicated in bold. Remark that several categories have zero count. The sample size of the completers decreases in consecutive survey parts because of the conditional approach excluding participants that had already dropped out.

3. *Supplementary tables*

Suppl. Table 1. Comparison between training and test set. Chi square tests were performed on absolute counts. Subscripts indicate the degrees of freedom.

	Category	Frequency (%)		χ^2 test
		Training	Test	
Gender	Male	85.51	84.93	χ^2_1=0.028 p=0.867
	Female	14.49	15.07	
Phase of study	1st bachelor	33.17	31.38	χ^2_4=3.629 p=0.459
	2nd bachelor	23.73	22.28	
	3rd bachelor	23.73	23.09	
	Transfer student	6.78	6.34	
	Master student	12.59	16.91	
Parental occupation	At least one engineering parent	28.50	28.55	χ^2_1=1.7e-30 p=1
	Non-engineering parents	71.50	71.45	
Vocation interest	No idea	9.00	9.98	χ^2_7=2.167 p=0.950
	PL	15.42	16.01	
	CI	7.71	8.26	
	OE	9.00	8.61	
	PL+OE+CI	6.43	7.06	
	PL+OE	19.79	17.56	
	PL+CI	15.94	17.73	
	OE+CI	16.71	14.80	
Migration status	Migration background	7.16	7.23	χ^2_2=2.015 p=0.365
	Non-migration background	91.82	90.53	
	Other	1.02	2.24	
Engineering persistence	Persistent	34.55	39.02	χ^2_2=2.005 p=0.367
	Somethimes thinking about a non-engineering career	49.16	44.84	
	Thinking about a non-engineering career	16.29	16.14	

Suppl. Table 2. Logistic model results for the missingness mechanism. Significant effects are presented in bold. The first column includes the model parameters with the respective category between brackets. Reference categories; gender: male, phase: 1; parents: 1 (=engineering parent), Q1: 0, Q3-6:1, Q7: 0, Q9-12: 1. Variance inflation factors for the predictor variables were below 5. Probability of survey completion is modeled (0: dropout, 1: completed), indicating for instance that 'phase 3' and 'phase 4' students have a higher probability to complete part 1 compared to phase 1 students. Firth correction was applied in the model for part 3-5 because of data separation.

	Part 1			**Part 2**			**Part 3-5**		
	Estimate	Se	P-value	Estimate	Se	P-value	Estimate	Se	P-value
(Intercept)	7.810	1.371	0.000	3.137	2.434	0.198	-4.491	3.181	0.158
Gender (V)	0.443	0.418	0.289	0.287	0.451	0.525	-0.340	0.340	0.318
Phase 2	0.699	0.395	0.077	0.187	0.415	0.653	0.086	0.390	0.825
Phase 3	1.562	0.466	**0.001**	1.049	0.544	0.054	-0.718	0.481	0.136
Phase 4	1.295	0.641	**0.043**	0.780	0.890	0.380	-1.314	0.725	0.070
Phase 5	0.886	0.468	0.058	0.332	0.609	0.586	-0.296	0.695	0.671
Parents (2)	-0.240	0.337	0.476	0.243	0.339	0.473	-0.102	0.299	0.734
Q1 (1)	0.697	0.487	0.153	1.200	0.540	**0.026**	0.698	0.487	0.152
Q1 (2)	1.442	0.547	**0.008**	1.007	0.578	0.081	0.500	0.533	0.348
Q3 (2)				0.726	1.566	0.643	1.826	1.540	0.236
Q3 (3)				0.837	1.398	0.549	2.796	1.465	0.056
Q3 (4)				0.494	1.401	0.724	2.638	1.472	0.073
Q4 (2)				-1.823	1.561	0.243	-4.708	2.176	**0.031**
Q4 (3)				-1.962	1.561	0.209	-4.609	2.174	**0.034**
Q4 (4)				-1.489	1.744	0.393	-4.025	2.268	0.076
Q5 (2)				1.453	1.140	0.202	0.946	1.088	0.385
Q5 (3)				1.279	1.107	0.248	1.237	1.076	0.251
Q5 (4)				2.308	1.260	0.067	1.073	1.142	0.347
Q6 (2)				-0.847	0.507	0.095	-0.127	0.368	0.730
Q6 (3)				-1.127	0.614	0.066	-0.546	0.476	0.252
Q6 (4)				-1.386	0.951	0.145	-0.524	0.928	0.573
Q7 (1)							0.468	0.313	0.135
Q7 (2)							0.488	0.383	0.202
Q9 (2)							-0.488	1.403	0.728
Q9 (3)							0.471	1.419	0.740
Q9 (4)							1.697	1.484	0.253
Q10 (2)							-0.647	1.522	0.671
Q10 (3)							-0.577	1.540	0.708
Q10 (4)							-0.235	1.564	0.881
Q11 (2)							0.499	0.591	0.399
Q11 (3)							-0.020	0.618	0.974
Q11 (4)							0.232	0.779	0.766
Q12 (2)							-0.009	0.371	0.982
Q12 (3)							-0.051	0.455	0.912
Q12 (4)							-0.711	0.714	0.319

Suppl. Table 3A. Standardized factor loadings CFA model 1. Results were obtained using 20 multiply imputed datasets. Se: standard error. The first item for each construct is the anchor item.

	Estimate	Se	P-value
CE (career exploration)			
Q25	0.753	0.022	0.000
Q26	0.777	0.022	0.000
Q27	0.728	0.022	0.000
Q28	0.703	0.022	0.000
Q29	0.580	0.023	0.000
Q30	0.545	0.023	0.000
Q31	0.584	0.024	0.000
Q24	0.583	0.023	0.000
RFC (role-fit confidence)			
Q8	0.769	0.024	0.000
Q9	0.778	0.024	0.000
Q10	0.784	0.024	0.000
Q11	0.707	0.023	0.000
Q12	0.644	0.024	0.000
Q7	-0.629	0.027	0.000
CC (competence confidence)			
Q20	0.560	0.032	0.000
Q21	0.650	0.033	0.000
Q22	0.460	0.030	0.000
Q19	-0.625	0.035	0.000
PRA (professional roles awareness)			
Q5	0.741	0.045	0.000
Q4	0.469	0.037	0.000
Q3	0.578	0.038	0.000
CA (competence awareness)			
Q13	0.673	0.042	0.000
Q15	-0.640	0.039	0.000
Q16	-0.374	0.035	0.000
Q17	-0.375	0.035	0.000

Suppl. Table 3B. Modification indices CFA model 1. MI: modification index exceeding a value of 20, sEPC: standardized expected parameter change. ~~: residual correlation, =~: cross-loading.

	mi	sEPC
Q7 ~~ Q19	105.254	0.404
RFC =~ Q19	43.551	-0.485
Q16 ~~ Q17	39.244	0.297
Q11 ~~ Q12	30.833	0.240
Q27 ~~ Q28	24.915	0.207
CE =~ Q10	23.693	0.183
Q28 ~~ Q24	22.130	0.203

Suppl. Table 3C. Factor correlations CFA model 1. ~~ indicates the correlation of a construct with the others listed underneath. Se: standard error.

	Correlation	Se	P-value
CE ~~			
RFC	0.312	0.016	0.000
CC	0.456	0.028	0.000
PRA	0.294	0.030	0.000
CA	0.272	0.029	0.000
RFC ~~			
CC	0.512	0.031	0.000
PRA	0.428	0.034	0.000
CA	0.292	0.031	0.000
CC ~~			
RU	0.300	0.048	0.000
CA	0.576	0.052	0.000
PRA ~~			
CA	0.565	0.056	0.000

Suppl. Table 4A. Standardized factor loadings final CFA model (model 4). Results were obtained using 20 multiply imputed datasets. Se: standard error. The first item for each construct is the anchor item.

	Estimate	Se	P-value
CE (career exploration)			
Q25	0.758	0.022	0.000
Q26	0.782	0.022	0.000
Q27	0.734	0.022	0.000
Q28	0.667	0.023	0.000
Q29	0.584	0.023	0.000
Q30	0.550	0.024	0.000
Q31	0.587	0.024	0.000
Q24	0.532	0.026	0.000
RFC (role-fit confidence)			
Q8	0.774	0.024	0.000
Q9	0.785	0.024	0.000
Q10	0.790	0.024	0.000
Q11	0.713	0.024	0.000
Q12	0.648	0.024	0.000
Q7	-0.580	0.027	0.000
CC (competence confidence)			
Q20	0.588	0.033	0.000
Q21	0.679	0.035	0.000
Q22	0.484	0.031	0.000
Q19	-0.594	0.035	0.000
PRA (professional roles awareness)			
Q5	0.739	0.045	0.000
Q4	0.470	0.037	0.000
Q3	0.579	0.038	0.000
CA (competence awareness)			
Q13	0.670	0.042	0.000
Q15	-0.637	0.040	0.000
Q16	-0.301	0.035	0.000
Q17	-0.302	0.035	0.000

Suppl. Table 4B. Modification indices CFA model 4. MI: modification index exceeding a value of 20, sEPC: standardized expected parameter change. ~~: residual correlation, =~: cross-loading.

	mi	sEPC
Q27 ~~ Q28	36.409	0.248
Q11 ~~ Q12	28.599	0.234
CE =~ Q10	24.459	0.187

Suppl. Table 4C. Residual correlations and factor correlations CFA model 4. Upper part: residual correlations, lower part: factor correlations.

	Estimate	Se	P-value
Q7 ~~ Q19	0.405	0.041	0.000
Q16 ~~ Q17	0.292	0.045	0.000
Q28 ~~ Q24	0.199	0.043	0.000
CE ~~			
RFC	0.317	0.016	0.000
CC	0.464	0.028	0.000
PRA	0.298	0.030	0.000
CA	0.284	0.031	0.000
RFC ~~			
CC	0.437	0.030	0.000
PRA	0.429	0.034	0.000
CA	0.310	0.033	0.000
CC ~~			
PRA	0.299	0.048	0.000
CA	0.600	0.055	0.000
PRA ~~			
CA	0.605	0.060	0.000

Suppl. Table 5. Measurement model in the MIMIC model. Estimates represent loadings based on standardized factors only. Results were obtained from 20 multiply imputed datasets. ~~ represents a correlation.

	Estimate	P-value
CE (career exploration)		
Q25	0.827	0.000
Q26	0.879	0.000
Q27	0.783	0.000
Q28	0.729	0.000
Q29	0.677	0.000
Q30	0.614	0.000
Q31	0.630	0.000
Q24	0.561	0.000
RFC (role-fit confidence)		
Q8	0.813	0.000
Q9	0.839	0.000
Q10	0.824	0.000
Q11	0.739	0.000
Q12	0.671	0.000
Q7	-0.637	0.000
CC (competence confidence)		
Q20	0.590	0.000
Q21	0.685	0.000
Q22	0.495	0.000
Q19	-0.632	0.000
PRA (professional roles awareness)		
Q5	0.754	0.000
Q4	0.502	0.000
Q3	0.615	0.000
CA (competence awareness)		
Q13	0.697	0.000
Q15	-0.680	0.000
Q16	-0.320	0.000
Q17	-0.334	0.000

	Estimate	P-value
1. Residual correlations		
Q7~~Q19	0.358	0.000
Q16~~Q17	0.288	0.000
Q24~~Q28	0.159	0.001
2. Factor correlations		
CE ~~		
RFC	0.327	0.000
CC	0.410	0.000
PRA	0.221	0.000
CA	0.160	0.000
RFC ~~		
CC	0.421	0.000
PRA	0.382	0.000
CA	0.290	0.000
CC ~~		
PRA	0.254	0.000
CA	0.570	0.000
PRA ~~		
CA	0.572	0.000

Suppl. Table 6. Model results for path model. Estimates represent loadings based on standardized factors only. Results were obtained from 20 multiply imputed datasets. ~~ represents a correlation and ~ represents a regression.

	Estimate	P-value
CE (career exploration)		
Q25	0.759	0.000
Q26	0.785	0.000
Q27	0.734	0.000
Q28	0.669	0.000
Q29	0.591	0.000
Q30	0.554	0.000
Q31	0.594	0.000
Q24	0.532	0.000
RFC (role-fit confidence)		
Q8	0.771	0.000
Q9	0.788	0.000
Q10	0.789	0.000
Q11	0.713	0.000
Q12	0.646	0.000
Q7	-0.573	0.000
CC (competence confidence)		
Q20	0.585	0.000
Q21	0.667	0.000
Q22	0.477	0.000
Q19	-0.540	0.000
PRA (professional roles awareness)		
Q5	0.727	0.000
Q4	0.472	0.000
Q3	0.570	0.000
CA (competence awareness)		
Q13	0.668	0.000
Q15	-0.637	0.000
Q16	-0.312	0.000
Q17	-0.306	0.000

	Estimate	P-value
1. Correlations		
Q7~~Q19	0.402	0.000
Q16~~Q17	0.281	0.000
Q24~~Q28	0.199	0.000
CFC ~~ CC	0.320	0.000
PRA ~~ CA	0.565	0.000
2. Regressions		
RFC ~		
CE	0.220	0.000
PRA	0.308	0.000
CA	0.070	0.283
CC ~		
CE	0.358	0.000
PRA	-0.178	0.160
CA	0.616	0.000
PRA ~		
CE	0.292	0.000
CA ~		
CE	0.277	0.000

4. R-code

1. Descriptive statistics	Github/1.Descriptive_stats_thesis.R
2. Exploratory factor analysis	Github/2.Exploratory_factor_analysis.R
3. Confirmatory factor analysis	Github/3.Confirmatory_factor_analysis.R
4. MIMIC model	Github/4.MIMIC model.R
5. Path model	Github/5.Path model.R