ORIGINAL PAPER

Development and Validation of a Learning Quality Inventory for In-Company Training in VET (VET-LQI)

Svenja Böhn[1] · Viola Katharina Deutscher[1]

Abstract

Despite the importance of dual VET for economic growth and stability, internationally, systems struggle with quality assurance and quality improvement. In recent years, numerous research efforts have been made to identify and describe quality aspects in dual VET, especially with regard to in-company training conditions as perceived by apprentices. This has resulted in an outsized number of test instruments, comprising numerous scales and numerous items. The research presented here aims to assemble and organize all existing survey instruments in the VET context and to develop a comprehensive and validated questionnaire for dual VET that measures workplace characteristics (VET-LQI). For this purpose, 43 test instruments were identified and categorized using a qualitative meta-synthesis, and integrated into a general theoretical framework (Tynjälä 2013). The results of the meta-synthesis reveal diverse content areas as the current foci of VET quality research. The applicability of any existing survey is limited, as the majority of studies either focus on a small range of selected categories and/or do not report the validation results of their scales. Hence, a synthe-sized item pool was used. Short scales were extracted and tested in seven commercial training occupations in Germany, which covered all identified content areas of VET quality research. On the basis of item and factor analyses, 22 scales (containing 99 items) are identified that, taken together, satisfactorily reflect all common workplace characteristic scales in existent measurement instruments. The resulting instrument provides a broad collection of short scales reflecting the various foci of a longstanding and diverse research tradition, and will allow future researchers to analyze in-company training conditions more comprehensively, given limited opportunities and time re-sources for testing.

Keywords Apprentice questionnaire · Factor analysis · Qualitative meta-synthesis · Training quality · VET

✉ Svenja Böhn
   boehn@bwl.uni–mannheim.de

Extended author information available on the last page of the article

Introduction

Many countries around the globe have adopted a dual vocational education and training structure (e.g. Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, the Netherlands, Portugal, Sweden, Switzerland). The defining feature of this kind of vocational education and training (VET) is the idea of equipping apprentices with both practical and theoretical knowledge by combining company-based training programs provided by the private sector, with a school-based component, usually provided by the public sector, leading to qualifications in nationally recognized occupations. Advocates claim that this kind of VET structure is particularly aligned to the needs of the labor market (Krekel and Walden 2016). Some studies even claim a positive effect on youth unemployment rates, as well as on economic competitiveness, growth and stability (e.g. EU 2018; Hanhart and Bossio 1998; Hanushek 2012; McIntosh 1999). However, despite their perceived importance for education and economic policy, VET actors internationally struggle with issues of quality assurance and quality improvement (Le Mouillour 2017). For instance, Negrini et al. (2015) have shown that training conditions vary widely in relation to different industries, companies and professions. Different training conditions might have lasting effects on VET outputs such as the acquisition of vocational competence, professional and personal development processes and, finally, the social integration of young adults. Therefore, quality research into the assurance of learning conditions might be the most important factor for successful implementation of dual VET, and a prerequisite to achieving its claimed benefits at the individual and economic levels. However, studies addressing VET learning conditions frequently reveal alarming results, especially in regard to the in-company training component: insufficient mentoring and instructions, insufficient feedback, inadequate cooperation with the learning venue, or a lack of equipment (e.g. Ebbinghaus et al. 2010; Virtanen and Tynjälä 2008) are just some issues raised. This is problematic because the workplace plays a crucial role in dual vocational training. Moreover, for many apprentices it reflects their first long-term contact with the working world (Rausch and Schley 2015).

Against this backdrop, there is growing interest in the identification and evaluation of key characteristics of in-company training conditions. To get insights into training companies, many researchers use a questionnaire design – despite the related disadvantages of retrospective surveys (Rausch 2013; Tourangeau 2000). In the vast majority of cases the focus is on the learner's perspective, and apprentice surveys are used (Böhn and Deutscher 2019; Rausch 2012). As Tynjälä (2013, p. 15) states, this is in line with a constructivist view of learning, where '(…) the presage factors do not affect the learning process directly(, but) rather through the learner's interpretation (e.g. Prosser and Trigwell 1999; von Glasersfeld 1995)'.

However, the findings of these different studies are only partially comparable; this impedes aggregation. A key problem is that many different features of in-company training conditions are operationalized, and priorities set, in different ways. Various test instruments for the operationalization of training conditions in the VET system do exist but unfortunately, they are only rarely coordinated. In more recent years, a large number of test instruments have been designed, redesigned and adapted. However, there is still a lack of adequate test instruments that would enable the reliable detection of training conditions in VET (Velten and Schnitzler 2012). Some of these new test

instruments are derived from test instruments used to measure workplace learning conditions in an employee context: for instance, the Job Diagnostic Survey (JDS) of Hackman and Oldham (1974) or the Work Design Questionnaire (WDQ) of Morgeson and Humphrey (2006). Others were especially designed for the VET context – for instance, the MIZEBA of Zimmermann et al. (1994) or the IBAQ of Velten and Schnitzler (2012). Existing test instruments differ in length and level of detail, but most importantly with regard to their respective content. Moreover, it is particularly problematic that studies only rarely report item and scale analysis results. Hence, the measurability, verifiability and ultimately the comparability of findings regarding in-company training conditions is impeded.

This context is the starting point for the research presented here. It was the aim of our research to assemble and organize all existing test instruments in the VET context, to give an overview of the breadth of VET learning quality research and, on the basis of this overview, to develop a validated questionnaire to measure the workplace characteristics of dual VET. The VET-learning quality inventory (VET-LQI) reflects the substantive heterogeneity of existing test instruments in various content areas in dual VET contexts, and offers validated single items and short scales for these content areas. The distinctive feature of the test instrument presented here is that it summarizes and reflects previous research efforts and can therefore be considered a rich potential source of information for measuring the quality of VET learning conditions. By offering short scales for diverse content areas it could serve future researchers to analyze in-company training conditions more comprehensively, given the limited testing time and to integrate and compare their results with respect to a general theoretical framework (Tynjälä 2013).

In order to realize this endeavor, we took a two-step study approach: First, we started by collecting and categorizing all existing apprentice survey instruments in the area of VET learning quality, then through a qualitative meta-synthesis generating an overview of the state of operationalization of VET learning quality research (study 1). As a second step, we then used the synthesized item pool to extract and test short scales for all identified content areas of VET quality research in a German context (study 2).

Theoretical Model: Quality of Workplace Learning in the Dual VET Context

In recent years, a growing research effort has been put into the identification and description of central quality aspects of learning in VET. The term 'quality' in the context of workplace learning in VET is however characterized by a lack of conceptual clarity (e.g. Van den Berghe 1997). Harvey and Green (1993) state that the conception of quality depends on perspective and usage and that the term can be used to describe processes as well as to evaluate results. Moreover, quality always means a comparison between conditions and the normative expectations of interest groups. It can then be interpreted as the extent of goal achievement (Mirbach 2009; Ott and Scheib 2002). From this point of view 'quality' can be basically operationalized. But as goal orientations vary with interest groups, reference to the respective goal perspective is necessary, and subsequently the scope of interpretation is thereby restricted.

Synthesizing the current state of research according to Klotz et al. (2017), we propose to define VET quality as the subjective perception of characteristics of vocational training that are conducive to certain outcomes. At this point, the key characteristics of Tynjälä's (2013) 3-p model – a general model to describe learning processes – are drawn upon: While different conceptions of the term prevail (e.g. Blom and Meyers 2003; Harvey and Green 1993), a broad consensus has emerged in the workplace learning community in regard to using a 'three pillar approach' when describing in-company vocational training quality aspects (e.g. Seyfried et al. 2000; Tynjälä 2013; Visser 1994). For instance, Tynjälä (2013) refers to Biggs' (1999) 3-p model by distinguishing presage, process and product factors, and combines this approach with a strong emphasis on the individual's – in this case the learner's – perspective (Fig. 1): The learner factors, as well as the learning context, are represented within the presage (input) dimension. The process dimension covers the description of workplace learning characteristics, including the structure and performance of work tasks or the learning individual's interaction with others. Finally, within the product (output) dimension, all learning related outcomes are summarized; these are mainly focused on the individual's personal and professional development (Tynjälä 2013).

While there is broad consensus about the distinction between input, process and output dimensions in general, the key characteristics specifying those dimensions vary greatly within existing studies. The absence of a common understanding regarding the specific content structure of input, process and output factors leads to a lack of conceptual clarity regarding VET quality and is reflected inter alia by the large number of test instruments that have been developed in this context that have little reference to each other.

As noted above, it was the aim of our research to assemble and organize all existing test instruments in the VET context and to develop a questionnaire measuring
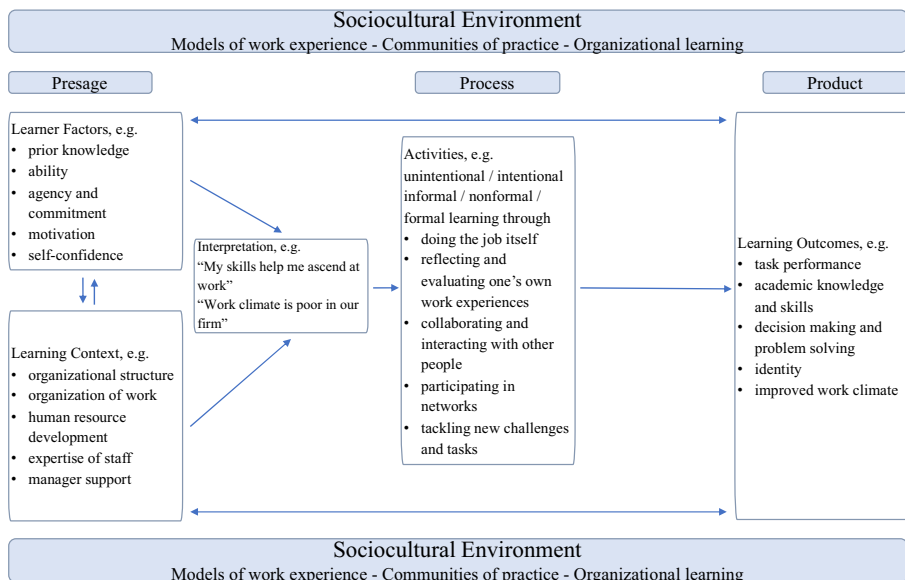

Fig. 1 The 3-p model of workplace learning (Tynjälä 2013, modified from Biggs 1999)

workplace characteristics in dual VET contexts. This focus is due to the fact that dual VET has some special characteristics, compared to workplace learning in general (Brooker and Butler 1997; Fuller and Unwin 2003; Raemdonck et al. 2014; Virtanen et al. 2009; Virtanen and Tynjälä 2008), especially in regard to three features: the role of the vocational school, curriculum-based work tasks, and training personnel. Usually, dual VET system conditions are governed by a legal framework. On the basis of this legal framework, learning is designed as a cooperative endeavor of vocational schools and private or public sector companies. As two organizational actors are integrated in the training process, there is a strong need for collaboration and coordination between the two learning venues. Hence, many apprentice surveys focus on aspects of learning venue cooperation (e.g. Brooker and Butler 1997; Dwyer et al. 1999; Ebner 1997; Feller 1995; Fink 2015; Heinemann et al. 2009; Keck et al. 1997; Nickolaus et al. 2015; Prenzel et al. 1996; Ulrich and Tuschke 1995; Virtanen et al. 2014; Walker et al. 2012). A second characteristic of dual VET systems relates to vocational curricula. Workplace learning is generally described as being mostly informal (Brooker and Butler 1997; Eraut 2004a, b; Marsick and Watkins 1990; Virtanen et al. 2009). In VET, on the other hand, the learning process of an apprentice at work is normally accompanied by a curriculum, structuring a domain in work tasks that have to be administered to apprentices in order to develop vocational competence. Moreover, as a third characteristic, dual VET is typically supported by training personnel who give instructions and ensure a certain level of formal organization. Further, because acquiring a formal occupational qualification is entailed, there is a need to assess the performance of an apprentice from time to time (Virtanen et al. 2009). This is why the design of work tasks and the role of the training personnel are both of high importance for in-company vocational training (Brooker and Butler 1997; Fuller and Unwin 2003; Virtanen et al. 2009). All three characteristics underline that workplace learning in the context of dual VET reflects a certain degree of formalization. Against this backdrop, apprentice surveys often focus strongly on questions related to those aspects (e.g. Beicht et al. 2009; Brooker and Butler 1997; DGB 2008; Ebner 1997; Ernst 2016; Feller 1995; Gebhardt et al. 2014; Heinemann et al. 2009; Hofmann et al. 2014; Keck et al. 1997; Koch 2016; Nickolaus et al. 2015; Prenzel et al. 1996; Velten and Schnitzler 2012; Virtanen et al. 2014; Zimmermann et al. 1994). As the goal of our research was to develop a comprehensive apprentice survey in the context of dual VET, we focused on instruments used in this context – though there is of course a significant thematic overlap to general workplace learning instruments. Further, the dual VET study is limited to learning at the workplace (in-company characteristics) rather than school-specific quality characteristics.

Study 1: Overview of Previous Apprentice Surveys and their Validity

Method

Given the large number of studies in this research area, their heterogeneous designs and conceptions of quality, a qualitative meta-synthesis (also referred to sometimes as qualitative meta-analysis), seemed particularly suitable for collecting and categorizing all existing apprentice survey instruments in the area of VET learning quality, as it allows the systematic and full integration of research results (Paterson 2012).

Qualitative meta-synthesis does however represent a rarely chosen form of analysis (Eisend 2014; Fricke and Treinies 1985; Glass et al. 1982). Following Lipsey and Wilson (2001), similarly to conducting a quantitative meta-analysis, one integrates the findings of several studies to develop an integrative overall result. In this case however, the database was qualitative in nature. For our study, the qualitative database consisted of the test items used in former apprentice surveys. In accordance with the process logic of meta-synthesis (e.g. Jensen and Allen 1996), the first two steps consist of (1) literature research and (2) literature selection. In our case a third step had to be added: this concerned dealing with the (3) collection of test instruments. The core of a qualitative meta-synthesis then, consists of the inductive determination of categories: (4) Item analysis and categorization. Finally, an integrative model of categories was built as a result of this methodological process.

Literature Research

For the VET context, a review of studies using apprentice surveys was conducted (Böhn and Deutscher 2019). This collection of studies, items and scales served as a starting point for the development of the VET-LQI (see study 2). In order to summarize the state of research into VET quality in apprentice surveys, we started out with a systematic literature search in eight databases (Business Source Premier, Deutscher Bildungsserver, EconLit, Education Resources Information Center (ERIC), Fachportal Pädagogik, Literaturdatenbank berufliche Bildung (LDBB), Social Sciences Citation Index (SSCI), Taylor & Francis), not differentiating in regard to publication type, profession, industry, language or country. The literature research included 21 German (Ausbildungsabbruch, Ausbildungsqualität, Ausbildungszufriedenheit, Berufsausbildung, betriebliche Ausbildung, betriebliche Ausbildungssituation, betriebliche Lernaufgaben, duale Ausbildung, duales System, Lehrling, Befragung, Fragebogen, Inventar) and English (Apprenticeship, On the job learning, VET, Vocational Education and Training, Workplace Learning, Work-based Vocational Training, Questionnaire, Survey) terms, which were combined in different ways. Thus, more than 13,000 search results were generated (including repeat results).

Literature Selection

After a detailed review and evaluation of the search results, all those studies were eliminated that (1) were only theoretically or conceptually founded, that used qualitative measurements or quantitative measurements but included no written survey (e.g. observation studies or interviews), (2) aimed at generating general assessments regarding the dual VET system without referring to the apprenticeship of the individual being questioned, (3) dealt with apprenticeship models that are not part of the dual system – meaning apprenticeships that do not integrate both a classroom based and a workplace based component and that therefore do not necessarily include practical experiences (including instructions and mentoring within a company), (4) focused solely on the classroom based instead of the workplace based component of the apprenticeship, (5) focused on a point of view other than the apprentices' – for instance, the perspective of training personnel or vocational teachers, (6) were written neither in English nor in German. On the basis of this literature search, in combination with the literature

selection criteria, 89 studies were deemed relevant. By retracing the references, an additional 23 studies were added.

Collection of Test Instruments

For a number of studies, the underlying test instrument was neither included in an appendix nor retrievable from supplemental materials or online resources. In such cases, we wrote to the authors asking for the test instruments to be provided. After waiting for a return time of eight weeks (return rate: 33.3%), in sum we had test instruments for 63.1% of all the studies. All those test instruments that were still not available after the expiry of the return time were not considered in the analysis. Similar applies to questionnaires that were incomplete or in a language other than English or German. In preliminary work for the analysis of test instruments, those questionnaires were identified that had been used in more than one study. This was necessary to avoid multiple analyses, which would have distorted the results. Finally, the literature research, selection and acquisition of test instruments provided the following results: (1) 112 studies were substantively relevant, but the underlying test instrument was available only for 71 of these. (2) One test instrument was available in two different versions, which were analyzed separately. (3) One study used a test instrument in a language neither English nor German. Therefore, it was not considered within the analysis. (4) In the case of three studies, only incomplete versions of the test instrument were provided. Therefore, they could not be considered within the analysis. (5) In the case of 37 studies we had no access to the test instruments at the expiry of the return time. (6) From those studies that provided the test instrument, fourteen were in English, the majority in German. (7) By checking multiple uses, 43 different test instruments could be extracted.[1]

Item Analysis and Categorization

The inductive determination of categories was conducted on the basis of Mayring's (2004) qualitative content analysis, including (1) generalization, (2) selection and (3) bundling, in order to connect and synthesize the test instrument items. First, all items were collected in tabular form, by excluding those questions dealing only with school-based aspects of the VET. Every single item was categorized separately on the basis of specific contents, and classified within a categorical system. Therefore, items with identical or similar contents were grouped: For example, the items 'Have you previously held a full-time job?' (NCVER 2000), 'Have you previously undertaken a pre-apprenticeship in the same industry area?' (Walker et al. 2012) or 'Do you have previous work experience?' (Virtanen and Tynjälä 2008) were grouped under the keyword 'personal background'. Subsequently, keywords were merged to categories, allowing a reasonable summary of contents. The grouping of items and the categorical summary were independent of positive or negative wording and different answer

---

[1] The item catalogue that served as a basis for designing VET-LQI can be accessed through this link: https://urldefense.proofpoint.com/v2/url?u=https-3A__www.bwl.uni-2Dmannheim.de_en_deutscher_research_&d=DwIFaQ&c=vh6FgFnduejNhPPD0fl_yRaSfZy8CWbWnIf4XJhSqx8&r=GJirMIpVRXTJ0ZCciTctcgrZyEVjA6OPFgTOHznqtEI&m=fHfTtXs6ifv2nA92ATWNZJdMlsDvPI_MQHPe9kZTrtM&s=wo3uhJ_XDt7c1GzsmdUSILcdnrtti2Nt5fCSlIjke6g&e= .

scales. It must be noted that the grouping of content-related items corresponded only to some extent, to descriptions and summaries within the scales of the original test instruments. Compared to the original studies, the results of this analysis deviate in part, either from the grouping of certain items in specific categories or the naming of certain categories. A codebook for the keywords and categories was developed. The entire analysis was carried out twice by one of the two researchers in tandem with a third person (intracoder reliability = .984, intercoder reliability = .926).

Results

Via this method, in total, 3356 items of the 43 test instruments were analyzed and classified into 30 categories in eight content groups. As we focused on workplace characteristics, a reduced version of the category system (covering 2343 items) was used (Fig. 2) – a supplement to Tynjälä's (2013) 3-p approach (Böhn and Deutscher 2019).

The results of this qualitative meta-synthesis served as a starting point for the subsequent analysis, and the development of an integrative test-instrument. As the aim was to reflect prior research efforts, first the user frequency of items and categories was determined – showing how many test instruments traced back to a certain category, and how many items could be assigned to each category, thereby referring to and extending the results of the qualitative approach to the analysis of test instruments identified by Böhn and Deutscher (2019). As a result, significant differences regarding the nominal and substantive design of the 43 test instruments were revealed (Table 1).

It can be noted that on average, less than 50% of all test instruments fell back to categories in the content area 'learning environment'. Those that did so mainly focused on the category 'usefulness of learning venue cooperation'; slightly more than half of all test instruments covered the 'vocational training framework'. There were big differences in the usage of categories from the content area 'work tasks', with a focus on 'overload'. Only one third of all test instruments, approximately, used questions
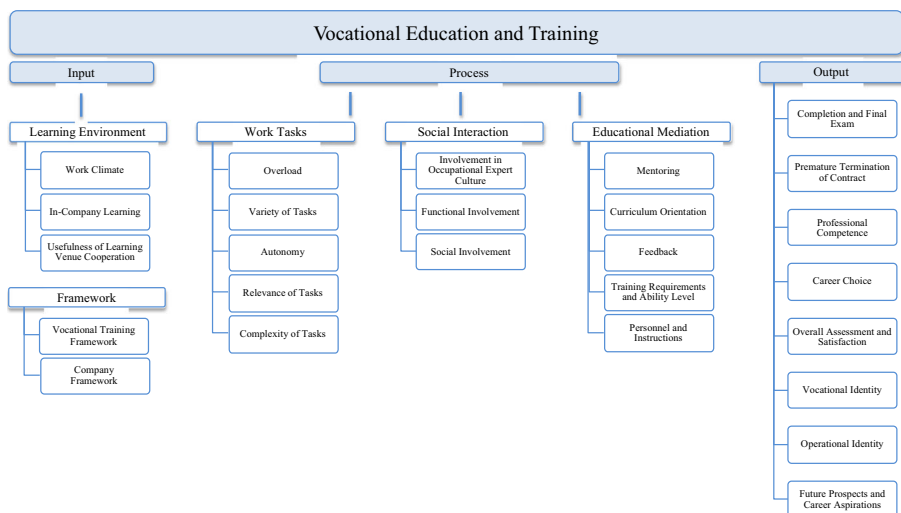

Fig. 2 Integrative category system based on a qualitative meta-synthesis of test instruments used in apprentice surveys in the context of dual VET

Table 1 User frequency and number of items per category, content area and dimension

Dimension	Content area	Category	A	B	C	D
Input	Learning Environment	Work Climate	17	39.5	42	1.8
		In-Company Learning	21	48.8	163	7.0
		Usefulness of Learning Venue Cooperation	24	55.8	113	4.8
	Framework	Vocational Training Framework	24	55.8	197	8.4
		Company Framework	25	58.1	77	3.3
Process	Work Tasks	Overload	22	51.2	162	6.9
		Variety of Tasks	17	39.5	43	1.8
		Autonomy	17	39.5	84	3.6
		Relevance of Tasks	15	34.9	35	1.5
		Complexity of Tasks	8	18.6	30	1.3
	Social Interaction	Involvement in Operational Expert Culture	13	30.2	27	1.2
		Functional Involvement	16	37.2	54	2.3
		Social Involvement	14	32.6	40	1.7
	Educational Mediation	Mentoring	23	53.5	59	2.5
		Curriculum Orientation	8	18.6	18	.8
		Feedback	22	51.2	93	4.0
		Training Requirements and Ability Level	22	51.2	47	2.0
		Personnel and Instructions	27	62.8	176	7.5
Output		Completion and Final Exam	11	25.6	24	1.0
		Premature Termination of Contract	8	18.6	12	.5
		Professional Competence	19	44.2	253	10.8
		Career Choice	25	58.1	134	5.7
		Overall Assessment and Satisfaction	35	81.4	180	7.7
		Vocational Identity	13	30.2	59	2.5
		Operational Identity	6	14.0	18	.8
		Future Prospects and Career Aspirations	27	62.8	203	8.7

A Number of test instruments with recourse to category ($N_t = 43$)

B Percentage of test instruments with recourse to category ($N_t = 43$)

C Number of items per category ($N_i = 2343$)

D Percentage of items per category ($N_i = 2343$)

related to the content area 'social interaction', with the focus being on the apprentice's 'functional involvement'. Test instruments differently reflected the content area 'pedagogical mediation', most of them concentrating on 'mentoring' and 'personnel and instructions'. VET outputs were also covered very differently, 'overall assessment and satisfaction' being the category chosen most often. Categories that covered by far the largest selection of items were 'vocational training framework', 'professional competence', 'overall assessment and satisfaction', and 'future prospects and career aspirations' (≥ 180 items each). There was much less variety with items in the content areas 'work tasks' and 'social interaction', and in some output categories.

Aside from their nominal and substantive heterogeneity, it is especially problematic that only a minority of studies reported validation results for their test instruments (Baethge-Kinsky et al. 2016; Dietzen et al. 2014; Fieger 2012; Hofmann et al. 2014; Lee and Polidano 2010; NCVER 2000; NCVER 2008; NCVER 2012; Nickolaus et al. 2009; Nickolaus et al. 2015; Prenzel and Drechsel 1996; Prenzel et al. 1996; Rausch 2012; Ulrich and Tuschke 1995; Velten and Schnitzler 2012; Velten et al. 2015; Virtanen and Tynjälä 2008; Virtanen et al. 2009, 2014; Wosnitza and Eugster 2001; Zimmermann et al. 1994, 1999). With only a few exceptions, the scales of these test instruments did however show satisfactory to excellent levels of internal consistency (e.g. Fieger 2012; Rausch 2012), and the factor analyses that were carried out in some cases, did confirm the model fit (e.g. Zimmermann et al. 1994). As the quality aspects of some test instruments are described in several studies, only ten of 43 test instruments, each of which focuses on selected aspects of the general model, can be considered to have been psychometrically validated.

The following conclusions can thus be drawn: A large number of test instruments for the VET context already exist – theoretically, researchers could make use of more than 3000 items. However, the recourse to any existing survey is limited, as the majority of studies either focus on a small range of selected categories and/or do not report the validation results of their scales. Further, the lack of a comprehensive and reliable test instrument for the analysis of in-company training conditions in VET must be noted. A test instrument that does reflect all characteristics identified in the category system is presented in this paper hereafter. A distinctive feature of this questionnaire is that whenever possible it refers back to existing items and scales. It therefore reflects foci of a longstanding and diverse research tradition to analyze in-company training conditions in dual VET systems.

Study 2: Design and Validation of the VET-Learning Quality Inventory (VET-LQI)

In the following section the development and design of the VET-LQI are set out. A first version of the test instrument, with 166 items, was pretested in 2017 in three vocational schools in the German state of Baden-Wuerttemberg. The sample consisted of 393 apprentices, associated with 15 different commercial VETs. This pretest led to some helpful adjustments, especially in regard to item reduction and wording. In particular those items were removed that caused difficulties in understanding or that did not improve the reliability of their respective scale, while otherwise being dispensable in terms of representing content validity adequately. This was the starting point for a shorter version of the VET-LQI, with 139 items, which is presented hereafter. The original version of the test instrument was in German, and was tested in three other vocational schools in Baden-Wuerttemberg in spring 2018. The structure of the questionnaire, the data, the item and factor analysis, and the results are presented subsequently.

Structure of VET-LQI

The VET-LQI is a synthesized test instrument that, on the basis of previous research efforts, reflects learning relevant quality aspects, with a focus on the key characteristics

of in-company training conditions. Therefore, it covers aspects of those categories that were identified in the qualitative meta-synthesis described above. The construction of new scales relies on a combination of efficient and frequently used items, together with a balance of newly developed items. In cases where existing scales and items showed little validated development, unduly answer scales or occupation-specific formulations, either adaption to a standard format or the creation of new items was necessary. In sum, 82 items could be accepted in full or in essence (Table 4). All other items were (re-)formulated and designed on the basis of the category's underlying codebook and, as it was the aim to design a test instrument that can be used in different occupational contexts, in phrases referring to respective occupations. Whenever necessary, the formulation of items indicated an additional wording such as 'in my company' or 'in my department' to clarify that apprentices were asked to focus on the in-company component of VET, instead of school-based aspects. Some items also explicitly indicated whether a question related to either the skilled occupation or the training company. This differentiation was particularly obvious in those scales representing the categories 'premature termination of contract', 'career choice', 'vocational identity', 'operational identity' and 'future prospects and career aspirations'. With the exception of five scales ('demographical factors', 'biography', 'academic performance', 'application process' and 'company framework'), the wording of items enabled respondents to answer on a 7-point Likert scale (1 'totally disagree' – 7 'completely agree'). Additionally, respondents could refuse to reply to every single item by choosing 'I do not want to or cannot answer this'. The test instrument was completely anonymized. No information was gathered that would enable researchers to draw conclusions as to the identity of individual persons, school classes, vocational schools or training companies. The VET-LQI was presented in a paper-pencil mode and contained 139 items in 31 scales. They were all derived from the category system – except the scale 'academic performance', which is content-related but which was intentionally separated from other questions dealing with the 'personal background' of the apprentice (Fig. 2).

First, a German version of the test instrument was designed. Then, every single item was translated to English. This translation process was carried out by two researchers and rechecked by a native English speaker. Hence, the VET-LQI is available in two languages (see Appendix). The results of the test of the German version are presented below.

Data Collection and Sample

The estimated time for answering the questionnaire was 45 min; the majority of apprentices finished in 30 min. The sample consisted of 428 apprentices ($N = 427$ after data editing), aged between 16 and 37 (mean: 20.5). 233 female and 194 male apprentices were surveyed, around 50% of them being in their first year of training, another 30% in their second year and approximately 20% in their third year of training. Data were collected in seven commercial VETs; the distribution is given in Table 2.

Table 3 presents weighted values for the sample in comparison to the statistical population of the seven training occupations analyzed. For both groups, t-test and chi^2-tests showed no significant differences in regard to age ($p = .426$), gender ($p = .586$) and education ($p = .100$).

Table 2 Sample distribution

Skilled occupation	Respondents	
	Absolute	Percentage
Bank Clerk ('Bankkaufmann/−frau')	15	4%
Logistics Specialist ('Fachkraft für Lagerlogistik')	29	7%
Industrial Management Assistant ('Industriekaufmann/−frau')	103	24%
Management Assistant in Informatics ('Informatikkaufmann/−frau')	22	5%
Office Management Assistant ('Kaufmann/−frau für Büromanagement')	87	20%
Management Assistant in Wholesale and Foreign Trade ('Kaufmann/−frau im Groß- und Außenhandel')	29	7%
Retail Salesman ('Kaufmann/−frau im Einzelhandel')	142	33%
Total	427	100%

First Item Analysis Results

Turning to the item analysis results, it has to be noted that those scales covering framework conditions ('vocational training framework' and 'company framework'), as well as two output scales ('completion and final exam' and 'career choice') remained unconsidered. Their content could either be reformulated or complemented on a very individual basis – depending on the use and context of intended studies, especially regarding questions covering personal details of the respondent, and the vocational training framework conditions. Moreover, those scales that primarily represented formative rather than reflexive measurement scales were also excluded from further analysis (Bollen and Lennox 1991; Diamantopoulos and Winklhofer 2001). This was

Table 3 Sample representativeness

	Age	Gender	Education
	Average age	Percentage of female apprentices	Percentage of apprentices with (1) < secondary school certificate (2) secondary school certificate (3) higher school certificate
Sample	20.5	54.3	(1) 9.4 (2) 49.6 (3) 41.1
Statistical population	20.6	53.2	(1) 14.1 (2) 41.4 (3) 42.0

The values are weighted over all vocational groups. The statistical population refers to the number of apprentices in Germany in 2017, according to the Federal Institute for Vocational Education and Training (BIBB); gender values referring to all apprentices in training, age and education referring to those entering their training in 2017. The average population age was increased by .5 as data were collected half a year after the start of nationwide training

especially so for questions related to the framework conditions and the output scales indicated above, which were unlikely to meet the requirements of factor analysis, as they did not describe reflexive theoretical concepts. For the remaining 99 items and 22 scales, a first aim of the item analysis was to shorten the questionnaire by identifying unsatisfactory items.

In the remaining 22 scales, 15 items were eliminated after a first examination, six because of a combination of low discrimination power ($< .3$, Ebel and Frisbie 1986) and the potential to improve the internal reliability (items 028, 029, 066, 069, 087, and 139). Nine items were excluded either because of a low scale correlation or because they appeared problematic for respondents in respect of their content or wording, as reflected in a high ratio of missing values of $> 5\%$ (items 032, 044, 047, 055, 062, 082, 096, 128, and 135). The internal consistency of the majority of scales was appropriate; 19 scales even reached good or excellent values (Cronbach's alpha $> .7$) while two scales were at least acceptable (Cronbach's alpha $> .6$, DeVellis 1991; Nunnally 1978; Robinson et al. 1991). The 'relevance of tasks' scale rated particularly poorly (Cronbach's alpha: .447). This was caused by a low discrimination power on two out of three items in this scale (item 059, with a discrimination power of .289, and item 061r, with a discrimination power of .217). The 'training requirements and ability level' scale was just below a good level of internal consistency (Cronbach's alpha: .684). It contained an item with a low level of discrimination power (item 070, with a discrimination power of .291). The 'overall assessment and satisfaction' scale, with a Cronbach's alpha of .657, was also just below a good level. The internal consistency could be improved by excluding item 124, which had a low discrimination power of only .234. However, for content-related reasons item 124 should be maintained.

Summarizing the findings, by excluding four additional items (items 033, 052, 088 and 124) the internal consistency of the scales could be further improved. Moreover, it has to be stated that there were five other items that generated more than 5% – set as a critical threshold – of refused or missing answers (items 026 and 064 with 6%, item 090 with 9%, and items 091 and 129 with 8% missing values). Nevertheless, content-related reasons indicated maintaining them initially, while keeping those twelve critical items in mind. In sum, 84 out of 99 items were kept for further analyses. The results of the item analyses are presented in Table 4.

Confirmatory Factor Analysis and Adaption of VET-LQI

Based on the category system in Fig. 2, a confirmatory factor analysis (CFA) was performed using the software packages R and lavaan (Rosseel 2012). CFA assumes reflective models, meaning that changes in the hypothetical construct cause changes in the indicator variables. Hence, all analyses were based on factor loadings and correlations between indicators and factors (e.g. Anderson and Gerbing 1988). When reflective models are adopted, all indicator variables describing one hypothetical construct are assumed to be highly correlated.

For CFA, first the free parameters within the base model have to be identified (Rosseel 2012). In this case, there were 22 latent variables (factors), each with an

Table 4 Item analysis results

Scale	Item	Cronbach's α (item deleted)	Mean	Standard Deviation	Difficulty	Discrimination Power
Learning environment						
Work climate (α .787)	021 If necessary, the employees in my company support each other. [3]	.748	5.87	1.23	.846	.570
	022 There is a personal atmosphere within my company. [3]	.736	5.71	1.23	.831	.611
	023 There is a bad working atmosphere in my company. [3][R]	.717	5.33	1.67	.748	.654
	024 There is strong competition between employees in my company. [3][R]	.762	5.52	1.38	.792	.520
	025 Employees in my company are rigorously monitored and controlled. [3][R]	.770	4.90	1.54	.628	.500
In-company learning (α .787)	026 Workplace learning in my company is characterized by different teaching methods. [4][*]		3.96	1.63	.371	.534
	027 Workplace learning in my company is characterized by the usage of different materials and media. [18]		4.24	1.63	.454	.469
Usefulness of learning venue cooperation (α .812)	030 What I learn at vocational school is important for the daily work in my company. [1]	.679	3.75	1.54	.245	.668
	031 When managing work tasks in my company, I benefit from knowledge I accumulated during vocational school sessions. [5]	.642	3.94	1.44	.294	.758
	033 The in-company vocational training and the vocational school are well coordinated. [6]	.887	3.40	1.59	.208	.524
Work tasks						
Overload (α .846)	045 In my company I feel under pressure of time at work. [10]	.817	3.19	1.54	.177	.678
	046 In my company I quarrel with supervisors, training personnel, colleagues or other apprentices. [10]	.843	2.37	1.53	.098	.485
	048 In my company others interfere my work. [9]	.837	2.59	1.44	.092	.536
	049 I have problems 'recharging batteries' in my spare time after work. [11]	.818	3.08	1.75	.189	.635
	050 Because of the daily demands in my company I feel totally exhausted, tired and drained. [11]	.801	3.82	1.71	.312	.740
	051 I often think 'I can't go on any longer'.	.804	2.88	1.81	.193	.716
Variety of tasks (α .824)	052 In my company I deal with a variety of work tasks. [12]	.902	5.09	1.39	.659	.516

Table 4 (continued)

Scale	Item	Cronbach's α (item deleted)	Mean	Standard Deviation	Difficulty	Discrimination Power
	053 In my company I work on new tasks every now and then. [12]	.647	4.53	1.60	.517	.781
	054 In my company work tasks are highly diversified. [12]	.662	4.42	1.64	.474	.767
Autonomy (α .872)	056 In my company I am flexible in the timing of work tasks. [12]	.853	4.18	1.53	.388	.774
	057 In my company I am able to decide on my own what means to take to reach a goal. [12]	.817	4.32	1.43	.430	.787
	058 In my company I am given an enormous amount of freedom in doing my job. [12]	.788	4.35	1.43	.450	.773
Relevance of tasks (α .447)	059 In my company I am confronted with responsible tasks. [3]	.428	5.41	2.76	.708	.289
	060 In my company I work on 'real tasks'. [8]	.228	5.67	1.27	.827	.443
	061 In my company I have to deal with several tasks that are not part of my vocational training program (e.g. make coffee, copying, etc.). [2][R]	.440	4.85	1.86	.619	.217
Complexity of tasks (α .841)	063 In my company work tasks are characterized by considering a wide range of information. [1]	.777	4.81	1.29	.574	.714
	064 In my company work tasks are characterized by considering a wide range of objectives. [1][*]	.756	4.56	1.22	.481	.665
	065 In my company work tasks are characterized by considering changes over time. [1]	.805	4.99	1.33	.625	.650
Social interaction						
Involvement in operational expert culture (α .881)	072 I am involved in the improvement of work processes in my company. [4]	.830	4.24	1.72	.441	.772
	073 My ideas and proposals are considered in my company. [4]	.813	4.43	1.64	.507	.791
	074 I am involved in the discussion of technical and professional issues in my company. [19]	.853	3.94	1.73	.376	.747
Functional involvement (α .801)	075 Continuous collaboration is part of the daily work in my company. [12]	.769	5.48	1.29	.781	.568
	076 My work tasks are characterized by a close coordination with other employees in my company. [20]	.794	4.97	1.43	.588	.486
	077 I am involved in all work tasks in my department. [4]	.754	4.81	1.63	.577	.618
	078 Basically, my work tasks play a crucial role for my department. [3]	.751	5.32	1.42	.708	.624

Table 4 (continued)

Scale	Item	Cronbach's α (item deleted)	Mean	Standard Deviation	Difficulty	Discrimination Power
Social involvement (α .786)	079 I am well integrated into the operational working procedures. [7]	.747	5.46	1.30	.764	.646
	080 Employees in my company are interested in me. [21]	.680	5.45	1.25	.770	.640
	081 Employees in my company are interested in my private well-being. [13]	.750	4.84	1.71	.620	.599
	083 Employees in my company seem disturbed by my presence. [3][R]	.758	6.27	1.14	.916	.506
	084 Employees in my company ignore me. [3][R]	.747	6.37	1.13	.921	.537
Mentoring (α .759)	085 In my company nobody feels responsible for me. [6][R]	.653	5.43	1.65	.751	.602
	086 In my company I am completely left alone to myself. [3][R]	.558	5.31	1.56	.702	.552
	088 When I ask training personnel or colleagues for help they immediately support me. [21]	.777	5.79	1.30	.842	.455
Curriculum orientation (α .824)	089 I do know my in-company training plan. [15]	.787	5.03	1.76	.658	.650
	090 The arrangements of my in-company training plan are observed. [15][*]	.701	4.77	1.69	.593	.737
	091 My in-company training program is implemented without a formal training plan. [22][R][*]	.782	5.18	1.92	.663	.661
Feedback (α .721)	092 In my company good performances are not praised. [6][R]	.678	5.09	1.75	.662	.496
	093 Normally I do know whether I perform work tasks satisfactorily or not. [12]	.677	5.26	1.21	.715	.487
	094 I find it hard to figure out whether I perform work tasks satisfactorily or not. [12][R]	.674	5.25	1.44	.691	.496
	095 The training personnel and my colleagues let me know whether I perform work tasks satisfactorily or not. [23]	.603	4.94	1.48	.633	.600
Training requirements and ability level (α .684)	067 In my company I am confronted with tasks that are too complicated. [14]	.538	2.73	1.21	.071	.484
	068 In my company I am confronted with tasks I am insufficiently trained and prepared for.	.632	3.06	1.42	.137	.375
	070 In my company I am confronted with tasks that are challenging for me. [3]	.677	3.83	1.40	.286	.291
	071 In my company I am confronted with easy tasks. [22][R]	.624	3.46	1.24	.153	.477
Personnel and instructions (α .908)	097 Those who train me on the job are able to answer difficult technical questions. [13]	.898	5.46	1.42	.788	.732
	098 Those who train me on the job can explain well. [14]	.883	5.39	1.34	.765	.798

Table 4 (continued)

Scale	Item	Cronbach's α (item deleted)	Mean	Standard Deviation	Difficulty	Discrimination Power
	099 There is a lot I can learn from those who train me on the job. [1]	.872	5.56	1.37	.798	.832
	100 Those who train me on the job show that they enjoy their work. [16]	.898	4.90	1.57	.612	.729
	101 Those who train me on the job are technically competent [24]	.888	5.70	1.27	.831	.754
Output						
Premature termination of contract (α .893)	105 I often think about terminating my training program prematurely. [25]	.858	2.08	1.52	.081	.834
	106 I think about terminating my training program prematurely because of operational reasons (e.g. working hours, quarrels with colleagues, etc.). [23]	.869	1.93	1.48	.081	.752
	107 I think about terminating my training program prematurely because of occupational reasons (e.g. dissatisfaction with tasks, lack of interest in field of activity, etc.). [23]	.864	1.95	1.48	.090	.792
	108 I think about terminating my training program prematurely because of vocational school reasons (e.g. quarrels with teachers, dissatisfaction with learning content, etc.). [23]	.892	1.66	1.13	.038	.548
	109 I think about terminating my training program prematurely because of personal reasons (e.g. family reasons, health reasons, financial reasons, etc.). [23]	.889	1.66	1.17	.043	.578
	110 I already search for an alternative vocational training. [23]	.875	1.86	1.44	.081	.708
	111 I already search for an alternative training company. [23]	.885	1.83	1.43	.079	.620
Professional competence (α .777)	112 I do have a high level of expertise. [26]	.756	4.87	1.08	.606	.489
	113 I do have a high level of social skills (e.g. strong communication skills, team spirit, etc.). [27]	.758	5.67	1.06	.857	.479
	114 I am able to perform most of the tasks in my company independently and without help. [5]	.708	5.59	1.05	.846	.641
	115 I feel well prepared to cope with future challenges in my job. [17]	.685	5.52	1.10	.810	.701
	116 At the end of my training program I will be able to perform every job-related task in my domain. [27]	.769	5.36	1.40	.750	.493
	124 I am strongly satisfied with the training program in my vocational school. [28]	.850	4.74	1.49	.574	.234

Table 4 (continued)

Scale	Item	Cronbach's α (item deleted)	Mean	Standard Deviation	Difficulty	Discrimination Power
Overall assessment and satisfaction (α .657)	125 I am strongly satisfied with the training program in my company. [28]	.525	5.13	1.61	.675	.496
	126 I am strongly satisfied with my vocational training in general. [29]	.192	5.18	1.30	.691	.758
Vocational identity (α .897)	127 My skilled occupation and me match very well. [1]	.878	5.20	1.30	.729	.709
	129 My skilled occupation is part of my personality. [7] [*]	.872	4.43	1.60	.490	.715
	130 I totally assimilate to my skilled occupation. [1]	.855	4.81	1.45	.591	.699
	131 I like telling others about my skilled occupation. [7]	.860	4.78	1.69	.596	.691
Operational identity (α .910)	132 My training company and me match very well. [30]	.886	5.13	1.54	.686	.719
	133 I am proud to work for my training company. [31]	.873	4.98	1.67	.651	.735
	134 My training company is a place like home for me. [7]	.906	3.99	1.96	.447	.666
	136 I like telling others about my training company. [7]	.871	4.92	1.81	.655	.740
Future prospects and career aspirations (α .777)	137 After finishing my vocational training I would like to work in that occupation. [23]	.639	4.98	1.81	.641	.639
	138 After finishing my vocational training I would like to work in my training company. [23]		4.57	2.04	.549	.639

[R] Item reversed

[*] Proportion of missing or refused answers > 5%

Answer format: 1 'strongly disagree' – 2 'mostly disagree' – 3 'slightly disagree' – 4 'partly agree' – 5 'slightly agree' – 6 'mostly agree' – 7 'completely agree' / alternatively: 'I do not want to or cannot answer this'

Items that were accepted in full (only one, and in particular the oldest source is given, as it can be assumed that newer publications also refer to older items and scales): [1] Heinemann et al. 2009, [2] Ernst 1997, [3] Zimmermann et al. 1994, [4] Fink 2015, [5] Keck et al. 1997, [6] Feller 1995, [7] Klotz 2015, [8] Ebbinghaus et al. 2010, [9] Koch 2016, [10] Kutscha et al. 2009, [11] DGB 2015, [12] Rausch 2012, [13] Velten and Schnitzler 2012, [14] Hofmann et al. 2014, [15] DGB 2008, [16] Prenzel et al. 1996, [17] Lehmann et al. 2005

Items that were slightly reformulated: [18] Beicht et al. 2009, [19] Fink 2015, [20] Nickolaus et al. 2015, [21] Velten and Schnitzler 2012, [22] Kutscha et al. 2009, [23] Feller 1995, [24] Ebner 1997, [25] Gebhardt et al. 2014, [26] Lehmann et al. 2005, [27] NCVER 2000, [28] Ulrich and Tuschke 1995, [29] DGB 2015, [30] Heinemann et al. 2009, [31] Hofmann et al. 2014

Table 5 Global fit indices for CFA: Base model

Estimator: MLM	Exact fit index		Approximate fit index	Residual based fit index	Comparative fit indices	
Model	Chi²/df ratio	p value (Chi²)	RMSEA	SRMR	CFI	TLI
1. 22 factors (base model): 84 items [a]	4885.633/ 3171 = 1.541	.000	.044	.080	.844	.828
2. one factor solution: 84 items [b]	9344.611/ 3485 = 2.681	.000	.078	.096	.476	.463

[a] with covariances between factors

[b] correlations between factors set to one

individual number of observable variables (indicators): in total, 84 items. Hence, 84 factor loadings needed to be estimated, as well as 22 covariances between the factors. Additionally, the residual variances of the indicators, as well as the variances of the factors had to be estimated – another 106 free parameters. In total, there were 212 free parameters. However, as for each latent variable, the factor loading of one indicator

Table 6 Global fit indices for CFA: Different models in comparison

Estimator: MLM	Exact fit index		Approximate fit index	Residual based fit index	Comparative fit indices	
Model	Chi²/df ratio	p value (Chi²)	RMSEA	SRMR	CFI	TLI
3. 22 factors: 82 items [a] [1]	4537.283/ 3008 = 1.508	.000	.043	.079	.859	.845
4. 22 factors: 80 items [a] [2]	4293.427/ 2849 = 1.507	.000	.043	.074	.861	.846
5. 22 factors: 76 items [a] [3]	3887.370/ 2543 = 1.529	.000	.044	.071	.864	.848
6. hierarchical model with one higher-level factor [4]	4551.848/ 2752 = 1.654	.000	.049	.088	.818	.812

[a] with covariances between factors

[1] eliminated items: 059 and 124

[2] eliminated items: 070 and 133

[3] eliminated items: 021, 046, 099, and 129
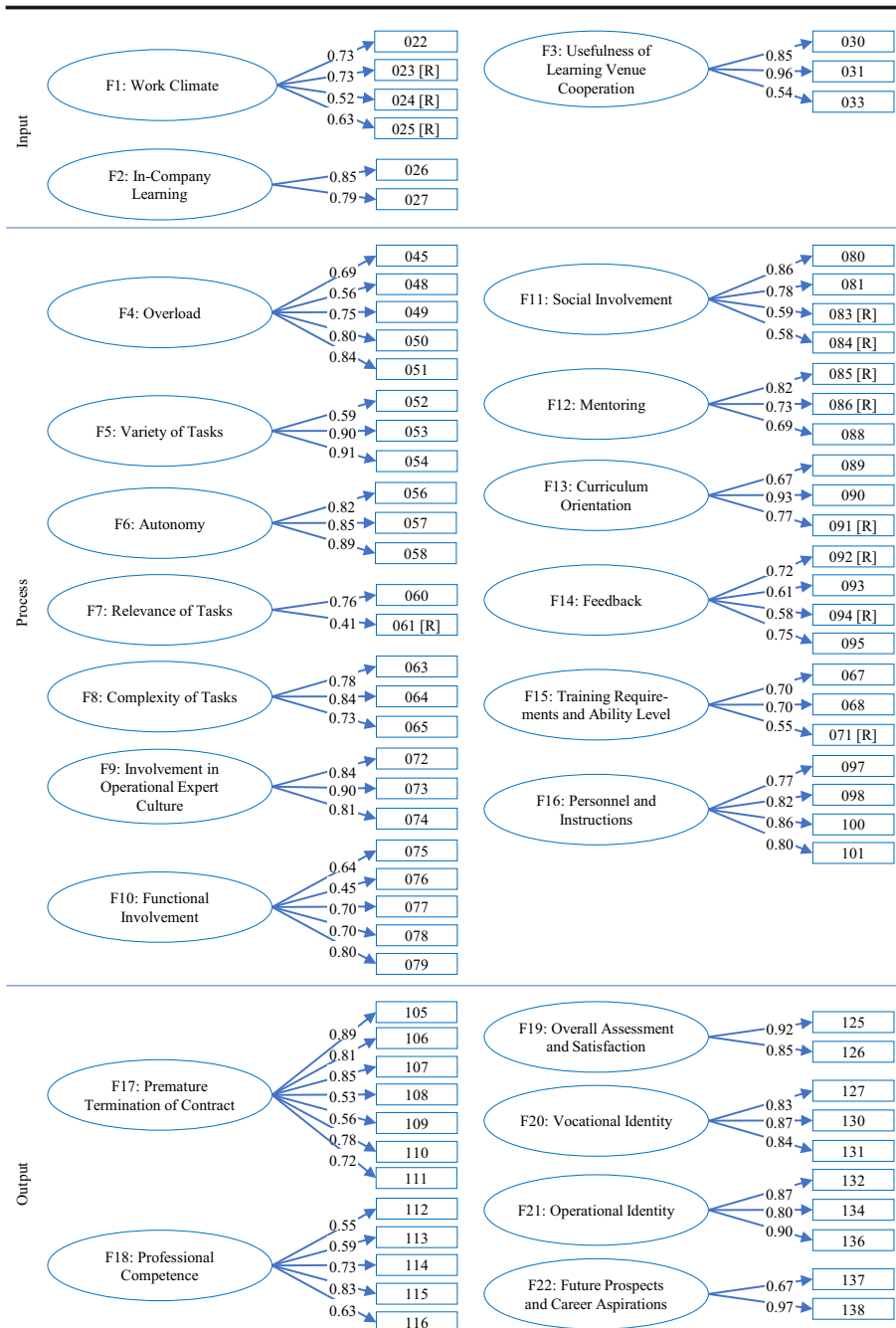
[4] eliminated items: see [1], [2] and [3]

Fig. 3 Item selection factor loadings based on model 5, referring to the 3-p model (Böhn and Deutscher 2019; Tynjälä 2013)

Table 7 Correlations between latent factors

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
F1	.662																			
F2	.450	.821																		
F3	.171	.378	.803																	
F4	-.590	-.320	-.171	.734																
F5	.588	.601	.219	-.472	.810															
F6	.431	.221	.139	-.401	.452	.851														
F7	.342	.276	.182	-.232	.538	.505	.611													
F8	.106	.260	.230	.082	.379	.225	.389	.782												
F9	-.047	.108	-.107	.480	.052	-.091	-.044	.302	.844											
F10	.581	.376	.235	-.391	.517	.658	.562	.262	-.123	.676										
F11	.559	.262	.191	-.341	.481	.514	.642	.273	-.262	.699	.713									
F12	.758	.345	.183	-.498	.547	.549	.651	.199	-.086	.664	.650	.749								
F13	.744	.407	.196	-.619	.559	.284	.468	.154	-.124	.415	.494	.684	.794							
F14	.461	.551	.278	-.459	.550	.170	.350	.159	-.087	.406	.341	.475	.735	.666						
F15	.636	.398	.317	-.553	.585	.442	.592	.188	-.170	.539	.548	.711	.733	.570	.650					
F16	.634	.476	.254	-.430	.509	.233	.392	.148	-.038	.355	.440	.514	.732	.491	.677	.818				
F17	-.624	-.359	-.145	.627	-.492	-.268	-.387	-.053	.140	-.369	-.405	-.537	-.589	-.488	-.534	-.512	.745			
F18	.245	.120	.127	-.320	.296	.388	.475	.140	-.431	.413	.544	.429	.191	.180	.386	.275	-.384	.675		
F19	.660	.488	.255		.582								.671	.616	.725	.625	-.584	.245	.886	
F20	.724	.507	.302	-.628	.622	.405	.493	.135	-.091	.546	.523	.677	.777	.682	.694	.714	-.738	.382	.906	.858
F21																				
F22																				

Diagonal values: square root of AVE

variable was set to one,[2] another 22 parameters were fixed. Hence, 190 free parameters had to be estimated. Second, we used MLM – the robust version of Maximum Likelihood (ML) – as an estimator for the purposes of conducting the CFA (e.g. Curran et al. 1996; Gold et al. 2003). MLM is based on the Satorra-Bentler scaled Chi^2 statistic, which yielded good results in robustness studies, especially when indicator variables deviated strongly from normality (Boomsma and Hoogland 2001; Chou and Bentler 1995; Schermelleh-Engel et al. 2003). Third, global goodness-of-fit indices for this model are reported on the basis of the recommendations by Schermelleh-Engel et al. (2003), as well as Matsunaga (2010), including Chi^2/df ratio, p value (Chi^2), RMSEA, SRMR, CFI and TLI (Bentler 1990; Bentler 1995; Bollen 1989; Browne and Cudeck 1993; Hu and Bentler 1995; Kaplan 2000; Matsunaga 2010; Schermelleh-Engel et al. 2003; Steiger 1990; Vandenberg 2006). However, it has to be noted that because of the high model complexity, due to the 22 latent variables, only RMSEA and SRMR can be considered interpretable, as their measures are insensitive to model complexity (Bentler 1995; Browne and Cudeck 1993; Kaplan 2000; Schermelleh-Engel et al. 2003; Steiger 1990). All other indicators deteriorate with sample size and the number of latent variables (Bentler 1990; Bollen 1989; Hu and Bentler 1995; Matsunaga 2010; Schermelleh-Engel et al. 2003; Vandenberg 2006). For the interpreting of RMSEA, the values reported by Browne and Cudeck (1993) were applied: RMSEA > .10 not acceptable, .08–.10 mediocre fit, .05–.08 acceptable fit and < .05 good fit. For the interpreting of SRMR, the values reported by Hu and Bentler (1995) were applied: SRMR < .10 acceptable fit, < .05 good fit.

Generally speaking, and following Matsunaga (2010, p. 108, citing to Kenny and McCoach 2003): '(I)t seems noteworthy that the number of items being analyzed in a given CFA is negatively associated with the model's goodness of fit. In other words, (...) the more the items, the worse the model fit'. Hence, not surprisingly, an examination of the local goodness-of-fit indices for the 22-factor solution indicated that improvements should be possible (Table 5). First – on an item-factor-level – single factor loadings should be relatively high and unambiguous, meaning that indicator variables with low factor loadings in general, or high factor loadings regarding more than one latent variable, might be problematic. Unfortunately, there is no exact threshold for defining 'low factor loadings' or 'high cross-loadings'. In this article, a threshold for single factors loadings is set to a minimum of | ± .4 |.[3] Cross-loadings appear problematic if differences in factor loadings are below | ± .2 |[4] (Matsunaga 2010). Second, significances of factor loadings should indicate that every indicator variable is affected by the latent variable within the population. Within this dataset, p values showed significant results for all items. Third, those indicator variables with low commonalities, or high values for uniqueness (indicating the proportion of variance

---

[2] The indicator variable that is set to one (the reference variable) is the variable identified as having the highest factor loading on each latent variable. For factor 1: item 023r, factor 2: item 026, factor 3: item 031, factor 4: item 050, factor 5: item 053, factor 6: item 058, factor 7: item 060, factor 8: item 064, factor 9: item 067, factor 10: item 073, factor 11: item 079, factor 12: item 080, factor 13: item 085r, factor 14: item 090, factor 15: item 095, factor 16: item 099, factor 17: item 105, factor 18: item 115, factor 19: item 126, factor 20: item 130, factor 21: item 133, factor 22: item 137.

[3] The following indicator variables had low factor loadings: items 059 and 124.

[4] The following indicator variables reached high factor loadings on more than one factor: items 021, 022, 046, 052, 060, 085r, 086r, 125, 133, 134, and 136.

that cannot be explained by the factors) should be treated with caution. For our purpose, the threshold was set to a value of above .6 for uniqueness, which applied to two variables.[5] Fourth, high correlations (> .8) between variables were checked. Fortunately, these were not found between variables of different factors; otherwise they could possibly have pointed to multicollinearity.

As the next step, several items were eliminated from the base model, owing to the following criteria: high uniqueness values (> .6), low factor loadings (< .4) and low discriminatory power (< .3). Three items appeared particularly problematic.[6] For content-related reasons, as two belonged to a common scale, initially only the worst item was eliminated. Hence, items 059 and 124 were excluded (Table 6, model 3). Then, two more items were eliminated due either to low discriminative power values or to high loadings on more than one factor[7] (model 4). Finally, another four items were excluded, for different reasons but with non-critical effects on the underlying scale in regard to number of items or reliability[8] (model 5).

Summarizing the findings, the differences in fit indices were evaluated, to compare the models. Even when applying a strict threshold of < 2 (Byrne 1991), the Chi2/df ratio indicated an acceptable fit for the base model, as well as all variations after item elimination – nevertheless, and not surprisingly, the Chi p value stayed significant in all cases. CFI and TLI did not reach an acceptable fit (Hu and Bentler 1999; Russell 2002); this is however expectable as, in contrast to RMSEA and SRMR, they punish model complexity, which was extensive for our estimated model. However, model comparison indicated an improvement for the non-hierarchical models after the elimination of eight items (see model 5 compared to models 1 to 4). RMSEA already indicated a good fit for the base model (model 1), and this improved slightly for models 3, 4 and 5. SRMR followed this logic, reaching its best value for model 5. Based on these fit indices, and on being the most parsimonious, model 5 was chosen to represent the best overall fit (Fig. 3). It has to be noted that fit indices also indicate an acceptable model fit when there is the assumption of a 22-factor solution with one higher-level factor (model 6).

Concerning the reliability of single items and scales, the following considerations should be noted: First, with the item selection in model 5, Cronbach's alpha decreased slightly in the case of seven scales.[9] As these changes in internal consistency were rather small, the aim of reducing the total number of items took precedence. Second, regarding the 'overall assessment and satisfaction' scale, item elimination caused an

---

[5] The following indicator variables reached high values of uniqueness: items 059 and 061r.

[6] The following indicator variables combine several critical values: items 059 (uniqueness: .71, factor loading: .372, discrimination power: .289), 061r (uniqueness: .70, factor loading: .404, discrimination power: .217), and 124 (uniqueness: .39, factor loading: .174, discrimination power: .234).

[7] The following indicator variables reached critical values in regard to one or more criteria: items 070 (discrimination power: .291) and 133 (correlation with item 132 > .8 and high factor loadings on two factors with a difference < | ± .2 |).

[8] Additionally, the following indicator variables were excluded: items 021 and 046, due to high factor loadings on more than one factor, 099, due to high correlations with item 098, and 129, due to the number of refused or missing answers being >5%.

[9] Change in Cronbach's alpha (compared to the original scale) for 'work climate' scale from .787 to .748, for 'overload' scale from .846 to .844, for 'relevance of tasks' scale from .447 to .428, for 'training requirements and ability level' scale from .684 to .677, for 'personnel and instructions' scale from .908 to .872, for 'vocational identity' scale from .897 to .872, and for 'operational identity' scale from .910 to .873.

improvement in the Cronbach's alpha.[10] In the case of the 'in-company learning' scale, item 027, due to its inconsistent wording, was split into two single questions: 'Workplace learning in my company is characterized by the usage of different materials' and 'Workplace learning in my company is characterized by the usage of different media'. Not only the structural validity of VET-LQI, but the convergent and discriminant validity also were checked. First, the average variance extracted (AVE) criterion, indicating convergent validity, yielded good results (> .5) for 16 of 22 factors, while 5 factors were just slightly below (> .4) (Fornell and Larcker 1981). Factor 7 – already performing poorly in reliability analysis – yielded an unsatisfactory value of .373. Discriminant validity, as the degree to which measures of different traits are unrelated, was assessed by the Fornell-Larcker criterion, as well as by analyzing the corrected correlations between the factors of the CFA. Thirteen factors met the Fornell-Larcker criterion (Fornell and Larcker 1981). The correlation matrix (Table 7) between the latent variables further explains that nineteen factors had moderate correlation values, but factors 19, 20, and 22 had high intercorrelations of > .8 (e.g. Evans 1996) that must be considered problematic.

Conclusion

Despite the possible disadvantages of retrospective surveys in general (Rausch 2013; Tourangeau 2000), the vast majority of researchers still use questionnaire designs to gain insights into quality aspects of VET. Moreover, previous research activities in VET have only partly focused on in-company training factors, primarily concentrating on the learners' factors. Particularly because of this focus and, additionally, the usage of different test instruments to operationalize training conditions, aggregating the findings regarding VET quality is still difficult. Therefore, the aim of this article was to present a validated selection of items and scales synthesizing the existing research regarding in-company training conditions that could be used in the context of dual VET. More than 3000 items were identified and categorized, and a selection were transferred into an integrative test instrument. This questionnaire may contribute to future research activities by providing a time-saving selection of validated items and scales for the analysis of in-company training conditions, and linking them to a general theoretical framework for vocational learning (Tynjälä 2013).

It was also our aim to present short scales that allow for the broad analysis of in-company training conditions, given the limited testing time. Those can, however, also be reused independently of one another in future research that aims to take a more specific focus. Especially in regard to the length of the questionnaire, a 5-point Likert scale could also be sufficient. Furthermore, it might be helpful to rename the middle category, to enable participants to express a neutral position ('uncertain' / 'unentschlossen' instead of 'partly agree' / 'trifft teilweise zu'). With the help of an item and factor analysis, useful indications on a parsimonious design of the test instrument could be gained. The total item number of 139 was reduced to 116 that collectively cover all the workplace characteristics scales existent in measurement instruments for dual VET. Analyses at both item and factor levels showed satisfactory results for the test instrument VET-LQI. All scales, except 'relevance of

---

[10] Change in Cronbach's alpha after elimination of item 124 from .657 to .850.

tasks', reached acceptable or good internal consistency values (.677 < Cronbach's alpha < .893). While reliability analysis, as well as convergent and discriminant validity analyses overall yielded satisfactory results, the 'relevance of tasks' factor also lacked convergent validity and should therefore be adapted for future use. In addition, with respect to discriminant validity, factor 19 ('overall assessment and satisfaction'), factor 20 ('vocational identity') and factor 22 ('future prospects and career aspirations') had high intercorrelations of > .8. Their combined use cannot be recommended, as this would likely cause empirical problems in future causal studies (e.g. multicollinearity). However, their high correlation seems theoretically plausible. This also explains why some factors did not meet the strict Fornell-Larcker criterion (factors 1, 7, 10, 12, 14, 15, 19, 20, and 22). Although these showed only negligible deviations of < .1, additionally, four of them contained at least one item with high cross loadings (factors 1, 7, 12, and 19). The combination of both indicators might be indeed problematic with regard to discriminant validity. Moreover, as our analyses were based on a limited sample size, had a regional (German) focus, and were confined to seven commercial occupations in the dual VET context, the instrument is yet to be validated for different samples and occupational fields, in future surveys, and of course with the use of the English translations of the scales. If necessary, occupation-specific adaptions and additions of items would be helpful in respect of item formulation, so that apprentices can more easily relate to the instrument. Our results are further limited by the fact that a test of other hierarchical models as part of the factor analysis was not possible, as the sample size was not sufficient for this kind of analysis. Future studies might show whether higher-order solutions yield superior fit-statistics.

Conceivably, the instrument could be adapted for longitudinal use, by dividing it according to the input, process and output dimensions and collecting data at the very beginning of VET, during VET (preferably multiple times) and at the very end of VET, or at least very shortly thereafter. The VET-LQI by its very design, reflects a process-oriented structure, particularly fruitful for such longitudinal research designs, where the aim is to identify relations between several inputs, processes and vocational training outputs. Such a longitudinal research design would then also allow for an assessment of measurement invariance. Moreover, some scales used within the VET-LQI could give an indication of how to design surveys focusing on the perspective of other actors in the VET context, e.g. training personnel or vocational teachers. Especially when different perspectives are to be compared – as has been the aim in at least some studies in the past (e.g. Krewerth et al. 2008; Pineda-Herrero et al. 2015; Saboga 2008; Walker et al. 2012) – this might be a worthwhile endeavor. Finally, it might also be possible to use the presented items outside of the traditional VET context, e.g. to survey dual students who are studying at universities of applied sciences and working part-time. Undoubtedly, certain items would then need to be adapted and tested again for validity purposes.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423.

Baethge-Kinsky, V., Baethge, M., & Lischewski, J. (2016). Bedingungen beruflicher Kompetenzentwicklung: Institutionelle und individuelle Kontextfaktoren (SiKoFak) [Conditions of competence development in vocational training: Institutional and individual context factors (SiKoFak)]. In K. Beck, M. Landenberger, & F. Oser (Eds.), *Technologiebasierte Kompetenzmessung in der beruflichen Bildung. Ergebnisse aus der BMBF-Förderinitiative ASCOT* (pp. 265–300). Bertelsmann: Bielefeld.

Beicht, U., Krewerth, A., Eberhard, V., & Granato, M. (2009). *Viel Licht – aber auch Schatten. Qualität dualer Berufsausbildung in Deutschland aus Sicht der Auszubildenden* (BIBB-report 09/09: Forschungs- und Arbeitsergebnisse aus dem Bundesinstitut für Berufsbildung) [Many bright spots - but shadows too. The quality of dual vocational training from the trainees' point of view]. Berlin: Bundesinstitut für Berufsbildung (BIBB).

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino: Multivariate Software.

Biggs, J. B. (1999). *Teaching for quality learning at university: What the student does*. Buckingham: Open University Press.

Blom, K., & Meyers, D. (2003). *Quality indicators in vocational education and training: International perspectives*. Adelaide: NCVER.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom in measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305–314.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation models: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 139–168). Chicago: Scientific Software International.

Böhn, S., & Deutscher, V. K. (2019). Betriebliche Ausbildungsbedingungen im dualen System – Eine qualitative Meta-Analyse zur Operationalisierung in Auszubildendenbefragungen [Training conditions in VET – a qualitative meta-synthesis for the operationalization in apprentice questionnaires]. *Zeitschrift für Pädagogische Psychologie, 33*(1), 49–70.

Brooker, R., & Butler, J. (1997). The learning context within the workplace: As perceived by apprentices and their workplace trainers. *Journal of Vocational Education & Training, 49*(4), 487–510.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park: Sage.

Byrne, B. M. (1991). The Maslach burnout inventory: Validating factorial structure and invariance across intermediate, secondary, and university educators. *Multivariate Behavioral Research, 26*(4), 583–605.

Chou, C. P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37–55). Thousand Oaks: Sage.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16–29.

Deutscher Gewerkschaftsbund (DGB). (2008). *Ausbildungsreport 2008 [training report 2008]*. Berlin: Deutscher Gewerkschaftsbund.

Deutscher Gewerkschaftsbund (DGB). (2015). *Ausbildungsreport 2015 [training report 2015]*. Berlin: Deutscher Gewerkschaftsbund.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Thousand Oaks: Sage.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research, 38*(2), 269–277.

Dietzen, A., Velten, S., Schnitzler, A., & Schwerin, C. (2014). *Einfluss der betrieblichen Ausbildungsqualität auf die Fachkompetenz in ausgewählten Berufen (Aqua.Kom). Abschlussbericht* [Effects of operational training quality on competence development in selected occupations (Aqua.Kom). Final report].

Dwyer, P., Harwood, A., Costin, G., Landy, M., Towsty, L., & Wyn, J. (1999). *Combined study and work paths in VET: Policy implications and analysis*. Adelaide: NCVER.

Ebbinghaus, M., Krewerth, A., Flemming, S., Beicht, U., Eberhard, V., & Granato, M. (2010). *BIBB-Forschungsverbund zur Ausbildungsqualität in Deutschland. Gemeinsamer Abschlussbericht zu den Forschungsprojekten 2.2.201 'Qualitätssicherung in der betrieblichen Berufsausbildung' und 2.2.202 'Ausbildung aus Sicht der Auszubildenden'* [BIBB research association with regard to training quality in Germany. Final report on 'quality assurance in vocational education and training' and 'vocational education and training from the apprentice's point of view']. Bonn: Bundesinstitut für Berufsbildung (BIBB).

Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of education measurement*. Englewood Cliffs: Prentice Hall.

Ebner, H. G. (1997). Die Sicht der Auszubildenden auf die Ausbildung [vocational education and training from an apprentice's point of view]. In D. Euler & P. F. E. Sloane (Eds.), *Duales system im Umbruch. Eine Bestandsaufnahme der Modernisierungsdebatte* (pp. 247–262). Centaurus-Verlagsgesellschaft: Pfaffenweiler.

Eisend, M. (2014). *Metaanalyse [Meta-analysis]*. München, Mehring: Hampp.

Eraut, M. (2004a). Informal learning in the workplace. *Studies in Continuing Education, 26*(2), 173–247.

Eraut, M. (2004b). Transfer of knowledge between education and workplace settings. In H. Rainbird, A. Fuller, & A. Munro (Eds.), *Workplace learning in context* (pp. 201–221). London: Routledge.

Ernst, C. (1997). *Berufswahl und Ausbildungsbeginn in Ost- und Westdeutschland. Eine empirisch-vergleichende Analyse in Bonn und Leipzig* [Career choice and training start in eastern and Western Germany. A comparative analysis in Bonn and Leipzig]. Bielefeld: Bertelsmann.

Ernst, C. (2016). *Forschungsprojekt 'Auszubildendenzufriedenheit'. Abschlussbericht* [Research project 'Apprentice satisfaction'. Final report]. Fakultät für Wirtschafts- und Rechtswissenschaften der Technischen Hochschule Köln.

EU (2018). *Addressing youth unemployment through outreach, activation and service integration*. ESF Youth Employment Thematic Network. Technical Dossier no. 9.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Belmont: Thomson Brooks/Cole Publishing.

Feller, G. (1995). Ansprüche und Wertungen junger Menschen in der Berufsausbildung. Wie und was Auszubildende (nicht) lernen wollen [Claims and evaluations of young people in vocational education and training. How and what apprentices do (not) want to learn]. *Berufsbildung in Wissenschaft und Praxis, 24*(2), 18–23.

Fieger, P. (2012). *Measuring student satisfaction from the student outcomes survey*. Melbourne: NCVER.

Fink, R. (2015). *Strategische Ausbildungsplanung und interne Evaluation: Steuerung pädagogischen Handelns, interne Evaluation, Fragebögen* [Strategic planning of vocational education and training and internal evaluation. Management of educational action, internal evaluation, questionnaires].

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50.

Fricke, R., & Treinies, G. (1985). *Einführung in die Metaanalyse [introduction into meta-analysis]*. Bern: Hans Huber.

Fuller, A., & Unwin, L. (2003). Fostering workplace learning: Looking through the lens of apprenticeship. *European Educational Research Journal, 2*(1), 41–55.

Gebhardt, A., Martínez Zaugg, Y., & Metzger, C. (2014). Motivationale, emotionale und selbstwirksamkeitsbezogene Dispositionen von Auszubildenden und deren Wahrnehmung der Lernumgebung und Lernbegleitung im betrieblichen Teil der beruflichen Grundbildung [Motivation, emotional and self-efficacy-related dispositions of trainees and their perceptions of the learning environment and learning support in the company part of initial vocational education and training]. *bwp@ Berufs- und Wirtschaftspädagogik – online, 26: Berufliche Bildungsprozesse aus der Perspektive der Lernenden*, 1–23.

Glass, G. V., McGaw, B., & Smith, M. L. (1982). *Meta-analysis in social research*. Beverly Hills: Sage.

Gold, M. S., Bentler, P. M., & Kim, K. H. (2003). A comparison of maximum-likelihood and asymptotically distribution-free methods of treating incomplete non-normal data. *Structural Equation Modeling: A Multidisciplinary Journal, 10*(1), 47–79.

Hackman, J. R., & Oldham, G. R. (1974). *The job diagnostic survey: An instrument for the diagnosis of jobs and the evaluation of job redesign projects*. Department of Administrative Sciences: Yale University.

Hanhart, S., & Bossio, S. (1998). Costs and benefits of dual apprenticeship: Lessons from the Swiss system. *International Labour Review, 137*(4), 483–500.

Hanushek, E. (2012). *Dual Education: Europe's secret recipe? CESifo forum 3/2012*. München: Ifo Institut.

Harvey, L., & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education, 18*(1), 9–34.

Heinemann, L., Maurer, A., & Rauner, F. (2009). *Engagement und Ausbildungsorganisation. Einstellungen Bremerhavener Auszubildender zu ihrem Beruf und ihrer Ausbildung. Eine Studie im Auftrag der Industrie- und Handelskammer Bremerhaven* [engagement and organization of vocational education and training. Attitudes of apprentices in Bremerhaven regarding their occupation and their training. A study of the German chamber of industry and commerce Bremerhaven].

Hofmann, C., Stalder, B. E., Tschan, F., & Häfeli, K. (2014). Support from teachers and trainers in vocational education and training: The pathways to career aspirations and further career development. *International Journal for Research in Vocational Education and Training, 1*(1), 1–20.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). London: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Jensen, L. A., & Allen, M. N. (1996). Meta-synthesis of qualitative findings. *Qualitative Health Research, 6*(4), 553–560.

Kaplan, D. (2000). *Structural equation modeling: Foundation and extensions*. Thousand Oaks: Sage.

Keck, A., Weymar, B., & Diepold, P. (1997). *Lernen an kaufmännischen Arbeitsplätzen. Berichte zur beruflichen Bildung* [Workplace learning in commercial occupations. Reports on vocational training], 199. Berlin und Bonn: Bundesinstitut für Berufsbildung (BIBB).

Kenny, D. A., & McCoach, D. B. (2003). Effects of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*, 333–351.

Klotz, V. K. (2015). *Diagnostik beruflicher Kompetenzentwicklung. Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne.* [Diagnosis of professional development. A didactic modeling for the commercial domain]. Wiesbaden: Springer.

Klotz, V. K., Rausch, A., Geigle, S., & Seifried, J. (2017). Ausbildungsqualität – Theoretische Modellierung und Analyse ausgewählter Befragungsinstrumente. In S. Matthäus, C. Aprea, D. Ifenthaler & J. Seifried (Eds.), *bwp@ Berufs- und Wirtschaftspädagogik online, Profil 5: Entwicklung, Evaluation und Qualitätsmanagement von beruflichem Lehren und Lernen. Digitale Festschrift für Hermann G. Ebner* [Development, evaluation and quality management of vocational teaching and learning. Digital Festschrift for Hermann G. Ebner] (pp. 1–16).

Koch, A. F. (2016). *Zufriedenheit in der Berufsausbildung. Konstruktion eines Fragebogens zur Erfassung intrinsischer Lern- und Arbeitsmotivation bei Auszubildenden im industriellen und kaufmännischen Sektor* [Satisfaction with vocational education and training. Construction of a questionnaire for the acquisition of intrinsic learning and working motivation among apprentices in the industrial and commercial sector]. Hamburg: Verlag Dr. Kovac.

Krekel, E. M., & Walden, G. (2016). Exportschlager Duales system der Berufsausbildung? [Export hit dual system of VET?] In L. Bellmann, G. Grözinger & W. Matiaske (Eds.), *Bildung in der Wissensgesellschaft* (pp. 55–70). Marburg: Metropolis-Verlag.

Krewerth, A., Eberhard, V., & Gei, J. (2008). *Merkmale guter Ausbildungspraxis. Ergebnisse des BIBB-Expertenmonitors.* [characteristics of good training practice. Results of the BIBB-Expertenmonitor]. Bonn: Bundesinstitut für Berufsbildung (BIBB).

Kutscha, G., Besener, A., & Debie, S. O. (2009). *Probleme der Auszubildenden in der Eingangsphase der Berufsausbildung im Einzelhandel – ProBE. Abschlussbericht und Materialien zum Forschungsprojekt an der Universität Duisburg-Essen* [Apprentices problems at the beginning of vocational education and training in the retail sector – ProBe. Final report and materials for the research project at the University of Duisburg-Essen].

Lee, W.-S., & Polidano, C. (2010). *Measuring the quality of VET using the student outcomes survey*. Melbourne: NCVER.

Lehmann, R. H., Ivanov, S., Hunger, S., & Gänsfuß, R. (2005). *ULME I. Untersuchung der Leistungen, Motivationen und Einstellungen zu Beginn der beruflichen Ausbildung* [ULME I. Analysis of performance, motivation and attitude at the beginning of vocational education and training]. Hamburg: Behörde für Bildung und Sport.

Le Mouillour, I. (2017). Duale Berufsbildungssystem in Europa vor ähnlichen Herausforderungen. Reformansätze in Österreich und Dänemark. [Dual VET systems in Europe facing similar challenges. Reform approaches in Austria and Denmark]. *Berufsbildung in Wissenschaft und Praxis, 3*, 37–38.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.

Marsick, V. J., & Watkins, K. E. (1990). *Informal and incidental learning in the workplace*. London: Routledge.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's, don'ts and how-to's. *International Journal of Psychological Research, 3*(1), 97–110.

Mayring, P. (2004). Qualitative content analysis. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A companion to qualitative research* (pp. 266–269). London: Sage.

McIntosh, S. (1999). *A cross-country comparison of the determinants of vocational training*. London: Centre for Economic Performance.

Mirbach, H. (2009). Qualität in der beruflichen Bildung – zur Auslegung des Qualitätsbegriffs [Quality of vocational training – interpretation of the quality concept]. In H. D. Münk & R. Weiß (Eds.), *Qualität in der beruflichen Bildung. Forschungsergebnisse und desiderate* (pp. 59–68). Bertelsmann: Bielefeld.

Morgeson, F. P., & Humphrey, S. E. (2006). The work design questionnaire (WDQ): Developing and validating a comprehensive measure for assessing job design and the nature of work. *Journal of Applied Psychology, 91*, 1321–1339.

NCVER (2000). *Student outcomes survey 2000*. National Report. Adelaide: NCVER.

NCVER. (2008). *Australian vocational education and training statistics. Student outcomes 2008*. Adelaide: NCVER.

NCVER. (2012). *Australian vocational education and training statistics. Student outcomes 2012*. Adelaide: NCVER.

Negrini, L., Forsblom, L., Schumann, S., & Gurtner, J.-L. (2015). Lehrvertragsauflösungen und die Rolle der betrieblichen Ausbildungsqualität [Premature termination of contract and the role of in-company training quality]. In K. Häfeli, M. Neuenschwander, & S. Schumann (Eds.), *Berufliche Passagen im Lebenslauf. Berufsbildungs- und Transitionsforschung in der Schweiz* (pp. 77–99). VS Verlag: Wiesbaden.

Nickolaus, R., Gschwendtner, T., & Geißel, B. (2009). Betriebliche Ausbildungsqualität und Kompetenzentwicklung [The quality of in-company education and training and the development of competence]. *bwp@ Berufs- und Wirtschaftspädagogik Online, 17*, 1–21.

Nickolaus, R., Nitzschke, A., Maier, A., Schnitzler, A., Velten, A., & Dietzen, A. (2015). Einflüsse schulischer und betrieblicher Ausbildungsqualitäten auf die Entwicklung des Fachwissens und die fachspezifische Problemlösekompetenz [The quality of in-company education and training and the development of competence]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 111*(3), 333–358.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

Ott, B., & Scheib, T. (2002). *Qualitäts- und Projektmanagement in der beruflichen Bildung. Einführung und Leitfaden für die Aus- und Weiterbildung* [Quality and project management in vocational training. Introduction and guideline for vocational education and training and further training]. Berlin: Cornelsen.

Paterson, B. L. (2012). 'It looks great but how do I know if it fits?': An introduction to meta-synthesis research. In K. Hannes & C. Lockwood (Eds.), *Synthesizing qualitative research: Choosing the right approach* (pp. 1–20). Oxford: Wiley.

Pineda-Herrero, P., Quesada-Pallarès, C., Espona-Barcons, B., & Mas-Torelló, Ó. (2015). How to measure the efficacy of VET workplace learning: The FET-WL model. *Education + Training, 57*(6), 602–622.

Prenzel, M., & Drechsel, B. (1996). Ein Jahr kaufmännische Erstausbildung: Veränderungen in Lernmotivation und Interesse [First year of vocational education and training in the commercial sector. Changes in learning motivation and interest]. *Unterrichtswissenschaft, 24*(3), 217–234.

Prenzel, M., Kristen, A., Dengler, P., Ettle, R., & Beer, T. (1996). Selbstbestimmt motiviertes und interessiertes Lernen in der kaufmännischen Erstausbildung [Self-determined, motivated and interested learning within the initial commercial training]. In K. Beck & H. Heid (Eds.), *Lehr-Lern-Prozesse in der kaufmännischen Erstausbildung: Wissenserwerb, Motivierungsgeschehen und Handlungskompetenzen* (pp. 110–127). Stuttgart: Steiner.

Prosser, M., & Trigwell, K. (1999). Understanding learning and teaching: The experience in higher education. Buckingham: SRHE and Open University Press.

Raemdonck, I., Gijbels, D., & van Groen, W. (2014). The influence of job characteristics and self-directed learning orientation on workplace learning. *International Journal of Training and Development, 18*(3), 188–203.

Rausch, A. (2012). *Skalen zu erlebnis- und lernförderlichen Merkmalen der Arbeitsaufgabe (ELMA). Forschungsbericht an der Otto-Friedrich-Universität Bamberg* [Scales on characteristcs enhancing experiences and learning (ELMA). Research report at the University of Bamberg].

Rausch, A. (2013). Task characteristics and learning potentials – Empirical results of three diary studies on workplace learning. *Vocations and Learning, 6*(1), 55–79.

Rausch, A., & Schley, T. (2015). Lern- und Motivationspotenziale von Arbeitsaufgaben als Qualitätsmerkmale des Lernorts Arbeitsplatz [Learning and motivation potentials of work tasks as quality aspects of workplace learning]. *Berufsbildung in Wissenschaft und Praxis, 44*(1), 10–13.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 1–16). San Diego: Academic Press.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36.

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in personality and social psychology bulletin. *Personality and Social Psychology Bulletin, 28,* 1629–1646.

Saboga, A. R. (2008). Level III apprenticeship in Portugal – Notes on a case study. *European Journal of Vocational Training, 45*(3), 121–143.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23–74.

Seyfried, E., Kohlmeyer, K., & Furth-Riedesser, R. (2000). *Supporting quality in vocational training through networking: CEDEFOP panorama.* Thessaloniki: European Centre for the Development of Vocational Training.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180.

Tourangeau, R. (2000). Remembering what happened. Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report. Implications for research and practice* (pp. 29–47). Mahwah, New Jersey, London: Lawrence Erlbaum.

Tynjälä, P. (2013). Toward a 3-p model of workplace learning: A literature review. *Vocations and Learning, 6,* 11–36.

Ulrich, J. G., & Tuschke, H. (1995). Probleme während der Lehre: Kritikpunkte von ostdeutschen Auszubildenden [Problems during apprenticeships: Critical points of apprentices in Eastern Germany]. *Sozialwissenschaften und Berufspraxis, 18*(3), 198–212.

Van den Berghe, W. (1997). *Indicators in perspective. The use of quality indicators in vocational education and training.* CEDEFOP document. Thessaloniki: European Centre for the Development of vocational training.

Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends. Where, pray tell, did they get this idea? *Organizational Research Methods, 9*(2), 194–201.

Velten, S., & Schnitzler, A. (2012). Inventar zur betrieblichen Ausbildungsqualität (IBAQ) [inventory of in-company training conditions (IBAQ)]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 108*(4), 511–527.

Velten, S., Schnitzler, A., & Dietzen, A. (2015). *Wie bewerten angehende Mechatroniker/−innen die Qualität ihrer betrieblichen Ausbildung? BIBB-Report 02/15: Forschungs- und Arbeitsergebnisse aus dem Bundesinstitut für Berufsbildung* [How do future mechatronics engineers evualuate the quality of their vocational education and training? BIBB-report 02/15: Research and work results of Bundesinstitut für Berufsbildung]. Berlin: Bundesinstitut für Berufsbildung (BIBB).

Virtanen, A., & Tynjälä, P. (2008). Students' experiences of workplace learning in Finnish VET. *European Journal of Vocational Training, 44*(2), 199–213.

Virtanen, A., Tynjälä, P., & Collin, K. (2009). Characteristics of workplace learning among Finnish vocational students. *Vocations and Learning, 2,* 153–175.

Virtanen, A., Tynjälä, P., & Eteläpelto, A. (2014). Factors promoting vocational students' learning at work: Study on student experiences. *Journal of Education and Work, 27*(1), 43–70.

Visser, K. (1994). *Systems and procedures of certification of qualifications in the Netherlands.* National Report. CEDEFOP Panorama. Berlin: European Centre for the Development of Vocational Training.

von Glasersfeld, E. (1995). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality.* How do we know what we believe to know? Contributions to constructivism (pp. 17–40). New York: Norton.

Walker, A., Smith, E., & Brennan-Kemmis, R. (2012). The psychological contract in apprenticeships and traineeships: Comparing the perceptions of employees and employers. *International Employment Relations Review, 18*(1), 66–81.

Wosnitza, M., & Eugster, B. (2001). MIZEBA – ein berufsfeldübergreifendes instrument zur Erfassung der betrieblichen Ausbildungssituation? Eine Validierung in der gewerblich-technischen Ausbildung [MIZEBA – a general instrument for the acquisition of in-company training conditions. A validation in the technical sector]. *Empirische Pädagogik, 15*(3), 411–427.

Zimmermann, M., Wild, K.-P., & Müller, W. (1994). *Entwicklung und Überprüfung des 'Mannheimer Inventars zur Erfassung betrieblicher Ausbildungssituationen (MIZEBA)'. Forschungsbericht der Universität Mannheim* [Development and validation of the 'Mannheim inventory for the acquisition of in-company training conditions (MIZEBA)'. Research report of the University of Mannheim].

Zimmermann, M., Wild, K.-P., & Müller, W. (1999). Das 'Mannheimer Inventar zur Erfassung betrieblicher Ausbildungssituationen' (MIZEBA) [The 'Mannheim inventory for the acquisition of in-company training conditions' (MIZEBA)]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 95*(3), 373–402.

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Svenja Böhn[1] · Viola Katharina Deutscher[1]

Viola Katharina Deutscher
deutscher@bwl.uni-mannheim.de

[1]   Economic and Business Education, Competency Development and Training Quality, University of Mannheim, L 4, 1, 68161 Mannheim, Germany