Constructing and validating authentic assessments: the case of a new technology-based assessment of economic literacy

Nina Charlotte Johanna Welsandt[1*] , Fabio Fortunati[2] , Esther Winther[2] and Hermann Josef Abs[1]

*Correspondence:
nina.welsandt@uni-due.de

[1] Educational Research and Schooling, Universität Duisburg-Essen, Universitätsstraße 2, 45141 Essen, Nordrhein-Westfalen, Deutschland
[2] Vocational Education and Training, Universität Duisburg-Essen, Universitätsstraße 2, 45141 Essen, Nordrhein-Westfalen, Deutschland

Abstract

Background: Authentic situations are considered a source of learning due to their real world relevance. This can encourage learners to acquire new knowledge. Increasing digitisation and associated resources, such as professional development opportunities for teachers, technology tools, or digital equipment for schools enable the development and implementation of authentic assessments. The basic academic principles for acquiring economic literacy are already provided in lower secondary school. This article examines, using the example of a new authentic technology-based assessment (TBA)—*Economic Literacy—Assessing the Status Quo in Grade 8* (ECON 2022) -, the processes involved in constructing a TBA. The purpose is to develop a curricular valid measurement instrument for surveying the current state of economic literacy in the 8th grade of a German federal state. This study explores which economic competencies students—typically between 14 and 15 years of age—possess in Grade 8, and what level of competence can therefore be expected of them at the beginning of a vocational training programme. The assessment is geared toward the curriculum of the subject of economics and is based on a domain model. This article presents the background and construction process for the development of ECON 2022 as a TBA.

Methods: To check the validity of test construction with a focus on the implementation of the authentic assessment and an analysis of difficulty-generating characteristics, the ECON 2022 test items were validated with an expert survey ($N = 25$). The two-stage data analysis comprised a descriptive quantifying analysis of the rating from the difficulty-generating characteristics specificity, cognitive demand and modelling and the design criterion authenticity. A set of experts rated the criteria. The expert survey was then compared with a previously conducted rating by the research team. The analysis of free-text comments on individual items was carried out discursively and qualitatively by the research team. Both sources of information were used to adapt the test items to measured item difficulties from the field test. For this purpose, items of great difficulty were changed to slightly easier items. In this context, the paper focuses on two central research questions: - How does the authenticity of a test environment relate to difficulty-generating criteria at item level? - Does the authenticity of a test environment have an impact on test results?

Results: Results are as follows. (1) The ECON 2022 assessment offers an example of a test design in which the use of TBAs can support innovative and interactive item development. (2) Using the ECON 2022 assessment enabled the implementation of an assessment of economic literacy using authentic situations and the mapping of different facets of economic competence. (3) The validation study showed that the actual item difficulty did not correlate significantly with the authenticity of the assessment, and authenticity thus did not contribute to item difficulty.

Conclusion: The results of the study show that we were successful in developing an authentic TBA in ECON 2022. ECON 2022 allows us to examine economic literacy in schools with a high degree of curricular validity and relevance and to analyse what level of competence and knowledge can be expected of students when they enter a vocational training occupation.

Keywords: Authentic assessment, Computer-based assessment, Technology-based assessment, Item design, Test design, Economic literacy

Introduction

Economic literacy is considered a component of general education which should be specifically promoted through the introduction of the subject of economics, which was launched in the federal state of North Rhine-Westphalia (NRW) in the school year of 2020/2021 (Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen [MSB] 2021). The launch of this new subject provided an occasion for developing a new knowledge test in the field of economic literacy. In this context, the project *Economic Literacy—Assessing the Status Quo in Grade 8* (ECON 2022) offered an opportunity to develop just such a new test.

Instruments in the field of education have changed substantially in recent years, nationally and internationally (Loerwald and Schnell 2016). In a detailed systematic review, Welsandt and Abs (2023) analysed 26 test instruments published between 1990 and 2020 with a total of 1124 items that measure competencies in economics across all age groups. The review showed that assessments differ considerably in their content and focus, and that they usually emphasise a particular aspect of the subject rather than covering all economic areas equally. Tests are aimed mainly at assessing a person's ability to recall factual information and are designed for adults as well as young people. Remarkably, the development of authentic assessments has not been a central focus even in recent times (Welsandt and Abs 2023). However, the ever-increasing potential of the technology-based assessment (TBA) to display images, videos, and audio sequences as part of the assessment offers new opportunities for making test environments authentic (Janesick 2006; Jude and Wirth 2007; Koh 2017). The major advantage of authentic test environments lies in dynamically designed test items that relate to real situations and that are based on skills relevant to everyday life (Janesick 2006). To incorporate these aspects, the goal of this paper is to focus on the innovative development of an authentic TBA for Grade 8 students in the field of economic literacy. The work presented in this paper is part of the research project ECON 2022.

To effectively measure the individual level of economic literacy through a test, it is crucial to first establish a clear definition of what constitutes economic literacy (Loerwald and Schnell 2016). In line with Beck (1989), This study defines economic literacy as a multidimensional construct with a linguistic-argumentative or mathematical-analytical

focus on the skills required to solve an economic problem. Test development was based on a domain model that has been derived from a systematic scientific and psychological analysis and tested for its curricular representativeness (Fortunati and Winther 2023a). The assumption was that authentic situations and the simulation of familiar behaviour would lead to an increase in the students' ability to use their economic skills to solve the test items. Therefore, individual items were embedded in an authentic economic narrative. The technical test environment of ECON 2022 was implemented via the CBA Item-Builder (Kröhne 2023).

In this context, there is a lack of research findings on assessments in economic education that associate student performance in achievement tests with the implementation of a computer-based authentic assessment. In this paper, we seek to address the question of how the authenticity of a test environment is related to possible difficulty-generating principles at item level. We also seek to determine to what extent test results are affected by the authenticity of the test environment. Clarifying the relations can help to identify and minimise possible biases in the test results. It is important to determine the impact of the implementation on authentic assessments to avoid disadvantaging any participant groups. A thorough review of this relationship will allow for the development of fair and balanced testing procedures that take appropriate account of the diversity of the test population.

To ensure the validity of the ECON 2022 assessment, an expert survey was conducted in addition to an analysis of field test data (Fortunati et al. 2024). In accordance with Beck (2020) and Sangmeister et al. (2018), the expert survey evaluated items based on three difficulty-generating design principles: (1) domain specificity, (2) cognitive demand level, and (3) item modelling (Klotz et al. 2015; Winther 2010). Furthermore, both authenticity and usability of the TBA (Sangmeister et al. 2018) were surveyed.

The paper is divided into five sections. "Introduction" section provides the introduction. "(Authentic) Assessments of Economic Literacy" section presents the current state of the art in test instruments for measuring economic literacy, with a specific focus on authentic computer-based design, and gives an overview of the principles of authentic testing according to Janesick and Gulikers. "Development of the ECON 2022 Assessment" section offers a detailed description of the development of the ECON 2022 assessment. The section begins with a theoretical and practical analysis of the design criteria for an authentic TBA, with a specific focus on how an authentic test environment can be implemented, and the importance of the lifeworld of the target group. Then the development of the ECON 2022 assessment is described with reference to the preceding considerations. "Validation and Revision of the ECON 2022 Assessment" section presents the expert validation of the ECON 2022 test items. Finally, "Discussion and Outlook" section discusses the results.

(Authentic) assessments of economic literacy

Tests of economic literacy

Everyday life is permeated by economic phenomena and problems that require economic literacy to comprehend and resolve. In the field of economic literacy, numerous test instruments have been developed in recent decades both nationally and internationally. These instruments often exhibit significant differences in terms of their conceptual

Welsandt *et al. Empirical Res Voc Ed Train* (2024) 16:4

Page 4 of 27

understanding of economic competence, test design, and target audience. Welsandt and Abs (2023) inspected which questions and decision options existing test environments raise for the development of future tests. For this purpose, a systematic review was conducted to examine the similarities and differences between the instruments for measuring economic literacy that have emerged over the past 30 years, and the extent to which the focus of these existing test instruments has evolved or changed. The population intervention comparison outcome (PICO) model was deployed to conduct a systematic search (Sayers 2007). The systematic review included all publications that used a measurement instrument or scale to assess basic economic literacy and reported on the original development, modification of a measurement instrument as a first step. Measurement instruments were excluded if they were not available in English or German, and if neither sample size nor Cronbach's alpha were reported. Altogether, 26 measurement instruments published between 1990 and 2020 were extracted; they included a total of 1124 items regarding economic literacy for all age groups. The analysis included survey format and technical implementation, year of publication, mode of implementation, response formats, content formats, the perspective of the economic subject dimension (Fortunati and Winther 2023a), the perspective of learning psychology (Marzano and Kendall 2007), and the perspective of authenticity (Janesick 2006). Table 1 lists the extracted measurement instruments that were included in the further analyses.

Overall, half of the test environments for measuring economic literacy included in the systematic review addressed children and young people under the age of 18. TBA was not a central feature of existing tests, and only 5 out of 13 test environments (38%) were computer based. With computer-based implementation, the focus was mostly on media support and on transferring tests from a paper-based to a computer-based format. The added value that the new setting might bring was not exploited. This was also evident in the choice of response formats, which were mostly limited to traditional single-choice, multiple-choice, or free-text answers (Welsandt and Abs 2023). Innovative answer formats, on the other hand, would include formats such as drag-and-drop items that allow test takers to move items physically around the screen to indicate their answer. Such formats can provide a more engaging and interactive experience for test takers while also gathering more detailed data for analysis, and this can be especially useful in tests that require spatial reasoning or sorting of items. Another innovative format would be the hotspot. In this format, test takers are presented with an image or diagram and asked to select a specific area by clicking on it. By moving a slider along a scale, test takers can select a value. This can be useful in tests that require numerical estimation or comparison. The innovative answer format of 'matching' presents two columns of items and asks the test taker to match them up. This can be useful in tests that require associations or pattern recognition. Lastly, there is also the possibility of more 'gamified' items. This format involves presenting test items in a game-like format such as quizzes, puzzles, or interactive simulations. This can be useful in engaging test takers and reducing test anxiety (Goldhammer and Kröhne 2020). Concept-mapping is another innovative format. This format can be implemented in less (create a map) or more restricted form (skeleton map). A concept map is a node-link diagram in which each node represents a concept, and each link identifies the relationship between the two concepts it connects (Schroeder et al. 2018). Test takers have to relate (given) concepts and label or choose

Table 1 Measurement instruments of economic literacy between 1990 and 2020 (based on Welsandt and Abs 2023)

Instrument	Country	Author, Publisher, or Institution	Year
Basic Economics Test (BET)	USA	Walstad, W. B.; Robson, D	1990
Economic literacy test *[Wirtschaftskundlicher Bildungs-Test (WBT)]*	Germany, Austria and Switzerland	Beck, K.; Krumm, V	1998
Test of Economic Knowledge (TEK)	USA	Walstad, William B.; Soper, John C	1998
Economic literacy of secondary school students *[Ökonomische Bildung von Schüler/innen Allgemeinbildender Höherer Schulen (OBHS)]*	Austria	Brandlmaier, E.; Frank, H.; Korunka, C.; Plessnig, A.; Schopf, C.; Tamegger, K	2005
Financial Fitness for Life Test (FFFT)	USA	Walstad, W. B.; Rebeck, K	2005
Test of Understanding in College Economics (TUCE)	USA	Walstad, W. B.; Watts, M.; Rebeck, K	2006
Jump$tart Coalition 2008 College survey (Jump$tart)	USA	Mandel, L Merrill Lynch Foundation	2008
National Assessment of Educational Progress (NAEP)	USA	National Center for Education Statistics	2012
The Financial Knowledge Scale (FKS)	USA	Knoll, M. A. Z.; Houts, C. R	2012
PISA 2012 Financial Literacy (PISA)	18 countries	OECD	2012
Financial education online: does it work? (FEO)	Italy	Nicolini, G	2012
A test of minimal economic knowledge in Germany (MEK)	Germany	Wobker, I.; Lehmann-Waffenschmidt, M.; Kenning, P.; Gigerenzer, G	2012
Test of Economic Literacy 4th Edition (TEL)	USA	Walstad, W. B.; Rebeck, K.; Butters, R	2013
Economic Competencies of Students Leaving Secondary School *[Ökonomische Kompetenzen von Maturandinnen und Maturanden (OEKOMA)]*	Switzerland	Schumann, S.; Eberle, F	2014
Impact of education on the financial literacy (IFL)	Slovakia, Czechia, Poland and Hungary	Tóth, M.; Lančarič, D.; Savov, R	2015
'Did you get it right?' flash cards. Maths in the area of basic financial education *[Stimmt's-Kärtchen. Rechnen im Bereich Finanzielle Grundbildung.]*	Germany	Deutsches Institut für Erwachsenenbildung	2015
CERAFORMA	Germany	Winther, E.; Klotz, V. K	2015
Test of Financial Literacy (TFL)	USA	Walstad, W. B.; Rebeck, K	2016
ALUSIM	Germany	Winther, E.; Seeber, S.; Weber, S.; Bley, S.; Festner, D.; Kreuzer, C.; Rudeloff, M.; Sangmeister, J.; Wiethe-Körpich, M	2016
PISA 2018 released financial literacy items (PISA)	20 countries	OECD	2018
The Financial Literacy Project (FLP)	USA	Breitbach, E.; Wagner, J	2018
National Financial Capability Study (NFCS)	USA	FINRA Investor Education Foundation	2018
Test of economic civic competence *[Leistungstest Wirtschaftsbürgerliche Kompetenz (WBK-T2)]*	Switzerland Germany	Ackermann, N	2018
Test of Economic Competence (TEC)	Germany	Kaiser, T.; Oberrauch, L.; Seeber, G	2019
Financial Literacy Test (FLT)	South Africa	Mudzingiri, C	2019

Welsandt *et al. Empirical Res Voc Ed Train*   (2024) 16:4

Page 6 of 27

Table 1 (continued)

Instrument	Country	Author, Publisher, or Institution	Year
Measurement of systems competence as an indicator of sustainable educational development *[Messung von Systemkompetenz als Indikator im Bereich Bildung für nachhaltige Entwicklung (SysKo-BNE)]*	Germany	Hartig, J.; Roczen, N	2020

a label for each link of two concepts. The systematic review showed that such formats were rarely used.

The systematic review highlighted that measurement instruments for economic literacy prioritised mapping one domain at a time rather than mapping all domains together. Despite the high relevance attributed to implementing lifeworld references in measurement instruments, the level of incorporation of authenticity in test formats seems inadequate at this time. Measurement instruments that are fully integrated into authentic settings remain the exception. For example, although all of the five computer-based test environments had items with a lifeworld reference, only two of the test environments (15%) were embedded in an authentic setting (Welsandt and Abs 2023).

Principles of authentic assessments according to Janesick and Gulikers

Authentic assessments consist of dynamic, real-life test items that are oriented towards abilities which are relevant to everyday life. Authentic problems are the origin of learning processes because of their strong connection with the lifeworld and because of their relevance, both of which motivate learners to gain new knowledge. Assessments should embed problems in authentic situations. The principle of the lifeworld reference increases the practical applicability for learners. Since learning tasks in the school context are always designed to be close to the real world at least in principle, it makes sense to implement this real world proximity in authentic assessments as well (Winther et al. 2022). Moreover, the didactic requirement of a test situation should ensure that the authentically conveyed learning content is queried in authentic test items (Klotz 2015). Authentic assessments require students to use their judgement to solve innovative items. Items require a specific set of student competencies in order to be solved. In authentic assessments, real-life situations are ideally simulated (Janesick 2006; Koh 2017). Gulikers et al. (2004) developed five dimensions to evaluate the level of authenticity in an assessment. At the task level, the degree of complexity should correspond to the level of responsibility of the natural work environment. This includes integrating knowledge, skills, and attitudes, as well as the complexity and relevance of the task for the learners. The physical environment simulated in the assessment should resemble the actual workplace environment. Computer-based implementation can help to increase authenticity. The assessment should reflect social relationships and processes in authentic professional settings. Furthermore, performance should be the primary basis of assessment and mirror the competencies that students would exhibit in real-life situations. Students should have multiple opportunities to demonstrate these attributes and capabilities through various tasks.

The assessment criteria should align with those applied in real workplace settings. The assessment's criteria and standards should be explicitly stated in order to ensure that students understand how their performance in a series of assessment tasks will be evaluated (Ersozlu et al. 2021).

Janesick (2006) has established six principles for authentic assessments. One, authentic assessments require students to demonstrate quality in performance or production, emphasising the significance of students' ability to apply knowledge effectively. Two, authentic assessments establish a strong connection between assessment tasks and the students' real-life experiences, ensuring relevance and practicality. Three, authentic assessments are characterised by their complex and multi-layered nature, requiring students to engage in diverse and interconnected tasks that mirror the complexity of real-life situations. Four, authentic assessments involve an ongoing process with multiple tasks. Five, authentic assessments seek to evaluate higher-order skills such as critical thinking, problem solving, and the application of knowledge in novel and meaningful ways. And six, complex feedback which is provided regularly plays a crucial role in authentic assessments because it allows students to self-adjust and improve their performance over time. By incorporating these principles, authentic assessments comprehensively analyse students' abilities and understanding beyond mere factual recall.

The development of ECON 2022 was based on principles 1–5. Various kinds of feedback can also be implemented as part of traditional individual assessment. Therefore, it doesn't appear as a necessary component of authentic assessment within the following analysis.

Authenticity has its origin in situated learning. Learning processes should be designed in such a way that the requirements they represent can be found in the real world, from which Winther (2010, p. 206) derived the requirement that test formats should also be authentic. Authentic assessments is directed to skills that are necessary for the lifeworld. These skills include the ability to solve problems, work independently, stay motivated, and regulate oneself while being aware of one's thought processes. Authentic assessments allow students to gain practical experience in using these specific skills and abilities, which are highly valued in the workforce (Villarroel et al. 2018, p. 2). Authenticity must be staged, which means the challenging situations must be modelled.

Designing and conducting authentic assessments also involves some challenges. For instance, implementation can lead to a considerable amount of extra work in test creation. Implementation requires time and financial resources and the acquisition of additional knowledge (Aziz et al. 2020, p. 763; Tanner 2001, p. 28). Depending on the design, authentic assessments produce an increased density of information and an increased processing effort because of their contextualisation in the target group's lifeworld. Moreover, the level of language competence required of students is often more complex in authentic test environments. For example, in authentic assessments students are often asked to explain how they solved mathematical items. Although this provides important insights into the students' understanding of mathematics, it also requires excellent language skills (Tanner 2001, p. 28). It is therefore conceivable that authentic assessments in and as of themselves can have their own difficulty-generating effect.

Against this background, this article focuses on the design of a technology-based test environment to provide an authentic assessment.

Development of the ECON 2022 assessment

Design criteria for an authentic technology-based assessment for economic literacy

Quality criteria in assessment construction

In this paper, the term 'assessment' essentially defines an instrument developed for collecting data about students' competences (Pellegrino et al. 2001). If assessment is understood as a process, it can include three steps: operationalising a valid construct, the actual testing, and interpreting test results (Klotz 2015, p. 68). Assessments vary with regard to multiple aspects, such as the mode of presentation, standardisation in stimulus materials, the response format, and the extent to which test materials are close to the test takers' lifeworld. Nonetheless, in all instances, tests need a standardised procedure for evaluating and scoring test takers' responses (AERA 2014, p. 2). The concept of reliability plays an essential role in the interpretation of test results. In this context, reliability refers to the consistency of results when a test procedure is administered several times, irrespective of the method or assessment (AERA 2014, p. 33). Validity is a further quality criterion, which is considered the most important aspect in developing and evaluating tests. According to the *Standards for Educational and Psychological Testing* (AERA 2014), validity is the question of the plausibility of the interpretation of test results; therefore, the main focus of this paper is interpreting the theoretical construct based on the test results (AERA 2014, p. 11). Furthermore, Mislevy and Riconscente (2005) identified two fundamental components of test construction that also had relevance for the creation of the ECON 2022 assessment: selecting test items with a clear reference to the aim of the assessment; and including reliability considerations. With tests in a school context, the aim of the assessment is therefore necessarily determined by curricula for general or vocational education. In addition, assessments should be oriented towards authentic, domain-typical learning and work processes.

Difficulty-generating design criteria

Concerning item construction, the three difficulty-generating criteria (specificity, cognition, and modelling) in the area of vocational assessments were followed to establish a connection between the requirements of occupation-specific action situations and cross-domain action situations of economic literacy and numeracy (Winther 2010). For the purposes of this study, a fourth design criterion was added, namely authenticity. Figure 1 below shows the decision trees of specificity, cognition, modelling and authenticity with the difficulty-generating criteria that differentiate between three difficulty levels.

Specificity is one of four criteria used for describing economic competencies, especially the difficulty of tests. In this paper, the understanding of specificity is built on the domain model (Fortunati and Winther 2023a), which, in turn, is based on subject-content theories. In line with Gelman and Greeno (1989), a distinction is made between domain-specific and domain-general content. The more specific the items are, the greater the requirement for comprehensive knowledge of economic concepts from multiple subdomains to solve the item. Conversely, domain-general items rely on generic knowledge and skill structures, which are prerequisites for tackling problem situations that bridge multiple domains. The transfer of general to domain-specific competencies can depend on the contextualisation. The findings of Hering et al. (2020, 2021), for
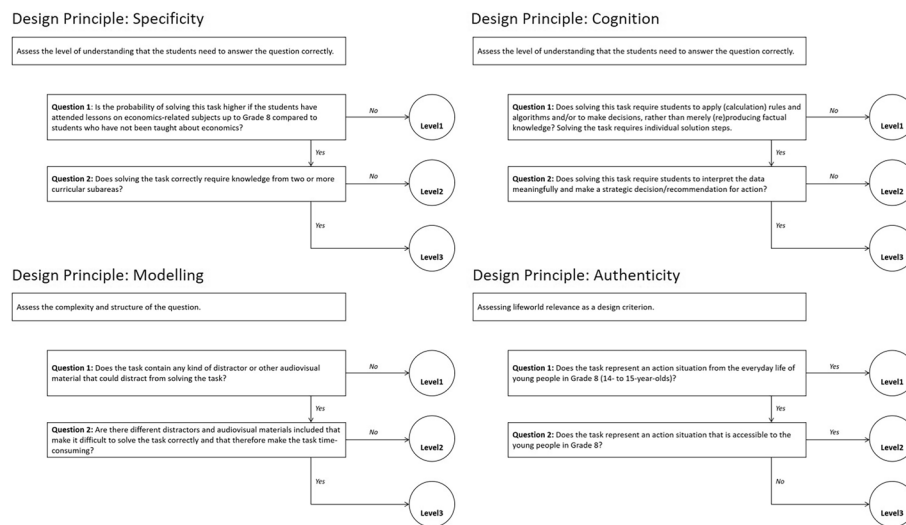
Fig. 1 Design principles

example, show that the transfer of general mathematical competencies into the context of commercial vocational training seems to depend in particular on the contextualization of the tasks. Regarding the probability of solving an item, levels 1 and 2 differentiate whether the teaching of economic knowledge was necessary or not. The items can be solved using general, economically relevant knowledge; they can also be solved at least partially without specific knowledge. Level 3, however, requires a combined knowledge of several economic subareas.

The second difficulty-generating characteristic is cognitive demand. The more cognitive resources required to process a test item, the more complex cognitive processes it engages. The taxonomies by Bloom et al. (1956) and Marzano and Kendall (2007) provided theoretical considerations regarding the level of cognitive demand of specific items and thus provided the basis for the decision tree on cognition. An item that can be assigned to level 1 can be solved solely by remembering and naming information; at level 1, knowledge only needs to be reproduced. A level 2 item requires information to be actively used, for example, by applying (calculation) rules or algorithms, or by making a decision. At level 3, data and results must be further interpreted and evaluated based on existing knowledge.

The modelling criterion represents the third difficulty-generating characteristic; it is based on cognitive load theory. It addresses the complexity of the presentation and perception of the item independent of the content difficulty of the item (Sweller et al. 1998). In other words, modelling seeks to measure artificial difficulties that occur independently of cognitive or content difficulty. For example, modelling features such as colour or presentation by means of audio, video, or continuous text could unintentionally influence the level of item difficulty. The decision tree for modelling focuses on the type and number of stimuli that might distract from a correct solution. If the approach is immediately obvious, it can be classified as level 1. If a distractor or audiovisual material is added that makes the item more difficult to solve, the item should be assigned to level 2. If several types of distractors and audiovisual material are used that could be misleading

or distracting, the item can be assigned to level 3. By analysing item modelling, it is possible to monitor artificially created difficulties that are detached from the cognitive and content-related difficulty (Klotz 2015).

The fourth design criterion is authenticity. Authenticity must be created, i.e. the challenging situation for measuring the economic competencies of the students must be modelled. It is assumed that situations that are familiar to young people from their everyday life make an item more accessible. If an item presents an action situation that is familiar to young persons from their everyday life, it can be classified as a level 1 item. If it is a situation that is accessible to young people at least in theory, it can be assigned to level 2. If the action situation cannot be expected to be accessible to young people, the item is assigned to level 3.

Development process of the ECON 2022 assessment

Target group: vocational education and training

In vocational and economic education, learning an occupation and the associated acquisition of competencies perform an important function for an individual's social integration (Beck et al. 1976). For vocational learning processes, school-based economic literacy education supports the development of area-specific competencies among trainees. The lifeworld environment shapes the competence acquisition process (Lempert 2009).

Recent school and curriculum reforms in the German federal state of NRW have tried to strengthen the vocational preparation of students. For example, from the 2014/2015 school year onwards, the initiative 'Kein Abschluss ohne Anschluss' [Guaranteeing next steps for school leavers: No school leaving certificate without subsequent opportunities for employment or qualification], mandated that all students in Grade 8 have to complete internships to explore occupational fields, which should be prepared and followed up at school (Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen 2020). Moreover, the introduction of the subject of economics focused the content of social science lessons much more strongly on vocational preparation. The ECON 2022 project took these initiatives as a starting point and targeted the additional part of the curriculum, which is specifically designed for vocational preparation and opening up connections with commercial training programmes.

Vocational education research has shown a clear predictive influence of domain-related economic competencies on the development of vocational competencies in commercial administrative professions (Achtenhagen and Winther 2008). Economic competence refers to the ability to navigate successfully situations that have economic implications, such as those related to the personal-financial, professional-entrepreneurial, and socioeconomic areas of life (Fortunati et al. 2024). This requires knowledge, skills, and abilities to understand and analyse economic problems in a specific context, develop solutions, make informed decisions, and reflect on actions taken. Previous studies on economic literacy focused on upper secondary school students and took a predominantly economic perspective in terms of content (Ackermann 2019). There is little empirical evidence regarding the development and structuring of economic literacy education at lower secondary schools as an important recruiting arena for commercial vocational training (Seeber et al. 2014). With that, the assessment developed in the context of the ECON 2022 project has been specifically designed to be conducted in

the preliminary stages of vocational training. The assessment helps to determine what business-related competencies students already have and what knowledge and skills can therefore be expected of them at the beginning of an apprenticeship.

Domain modelling and item construction

In constructing an assessment, ideas regarding a theoretical model and its output must be transferred into an appropriate assessment instrument. In addition to developing and compiling the items and the required materials, the measurement model must be reviewed, the scoring procedure prepared, and implementation techniques tested (Winther 2010). The curriculum-instruction-assessment triad (Pellegrino 2012) stipulates that test design should focus not only on valid test and item construction, but should also make continuous reference to the goals and content defined by the curriculum. The assessment must be tailored to the content of the school curriculum, which in turn is geared towards the learning fields. Accordingly, assessments must not only be coherent in themselves, but must also be meaningfully anchored within the entire education system, i.e. aligned with the curriculum (Klotz 2015, p. 48). According to Achtenhagen and Winther (2009), subject-didactic modelling of economic competencies is of great importance in constructing assessments, particularly as such modelling also addresses the subjects' process knowledge. To implement subject-didactic modelling, items should logically relate to each other in chronological order rather than depicting isolated partial aspects (Klotz 2015, p. 68). To construct a lifeworld reference for the target group, the knowledge, skills, and abilities of the target group must be recorded in authentically modelled situations.

Economic literacy is defined by Beck (1989) as a three-dimensional concept: (1) economic knowledge and cognition, (2) attitude towards economics, and (3) economics-related moral reflectiveness. Economic literacy is a prerequisite for achieving economic autonomy and participating in an evolving society. Individuals should be able to take part in society by developing their knowledge, skills, and abilities; they should understand and assess economic contexts which are located in the personal-financial, professional-entrepreneurial, and socioeconomic spheres of an individual's life, and make decisions (Beck 1989, p. 581). Knowledge in this context includes understanding fundamental elements of the economic world and learning about risks that can threaten economic well-being. Skills encompass generic cognitive processes within an economic context, such as information retrieval, comparison, extrapolation, and evaluation. They also entail fundamental mathematical and language abilities (OECD 2019). Skills can be understood as automatable action sequences that are performed routinely. Abilities comprise comprehensive mental tools with which a person can cope with challenges in particular situations. A routine cannot be applied but must be constructed for the situation. Competencies can be defined as complex combinations of abilities and skills which are the cognitive prerequisites for coping with specific lifeworld situations (Klieme et al. 2008; Hartig and Rauch 2008). Economic literacy encompasses all areas; though fundamental, pure knowledge or the execution of a routine are not sufficient. For example, in an arithmetic task, knowledge is needed, and routines can be performed.

Following Ackermann (2019), Fortunati and Winther (2023a) developed a domain model of economic literacy divided into three domains of life in which individuals are

confronted with economic situations. The personal-financial domain encompasses everyday economic conditions from the consumers' perspective and addresses the responsible management of personal finances. The professional-entrepreneurial domain contains economic challenges that individuals face in the workplace, which can be further categorised into general, occupational, and cross-occupational situations. The socioeconomic domain focuses on economic issues of high abstraction and complexity, often interconnected with political contexts. Furthermore, this domain concerns all citizens of a country (Fortunati et al. 2024). Sustainability is a cross-cutting dimension (Birindiba Batista et al. 2022) and is becoming increasingly important at both individual and corporate levels (Corsten and Roth 2012) from social and educational perspectives. It is being discussed in vocational (Haan et al. 2021; Rebmann and Schlömer 2020) and socioeconomic education (Schank and Lorch 2018) based on a holistic perspective. Education for sustainable development is a fixed curricular component and serves as a cross-sectional dimension with regard to economic content dimensions (KMK 2016). The ECON 2022 assessment focuses on sustainability as an overarching political concept, its implementation in the economic system, and its significance at the level of individual behaviour. In the questionnaire, individual attitudes are assessed that are not part of the test.

Technology-based assessments for measuring economic literacy and technical implementation in ECON 2022

The first TBAs were developed as early as the 1980s. Since then, the media used and the preparation of the content have changed. In addition to selecting the medium for delivery, the very design of the test environment and test items plays a much more important role (Steger 2019). For economic literacy, implementing a TBA is suitable for simulating an authentic lifeworld and realistic situations in which economic literacy applies (Winther and Achtenhagen 2009). A TBA is complemented by technology-based test construction and offers innovative possibilities for measuring knowledge, skills, and abilities. One aim of using TBAs for economic literacy can be to measure economic citizenship competencies, i.e. individuals' ability to understand and assess economic contexts and to form their own opinions based on their knowledge. The term TBA is a generic term for computer- and smartphone-based assessments (Steger 2019). Digital technologies such as laptops, tablets, and smartphones have become indispensable tools for competence measurement. They make it possible to collect data that goes beyond answering the items. Innovative use of digital technologies goes beyond simply digitising paper questionnaires in computer-based test formats and involves integrating multimedia elements or interactive tools. However, it must also be pointed out that so far, only a minority of existing tests have used digital technologies (Welsandt and Abs 2023). Therefore, a digital implementation of a test instrument is already innovative in and as of itself. In TBAs, more innovative answer formats can be used than in paper-based tests, and multimedia elements can be incorporated (Goldhammer et al. 2020). Furthermore, technology-based implementation enables the collection and analysis of processing data, i.e. data that allows conclusions to be drawn about item processing in addition to evaluating results data. Computer-based testing produces log data in log files (Goldhammer et al. 2020; Kögler et al. 2020). Analysing log data seems to be a suitable procedure for

inspecting the test takers' effort in processing the items. Furthermore, from a didactic perspective, TBAs open up relevant possibilities for the analysis of cognitive processing.

During the process of developing the assessment, work was carried out in parallel on the technical implementation and the construction of the content. Functions and answer formats for the technology-based test environment were defined. This made it possible to integrate additional help tools such as a virtual notepad, a calculator, and a help button to explain functions in the assessment. In addition to classic answer formats such as single-choice items, multiple-choice items, and free-text fields, more innovative formats were incorporated, including video sequences, sliders, drag-and-drop items that work by moving, combining, and placing different elements, or items in which incorrect answers must be crossed out. In the preliminary design stages, a suitable program for implementation was investigated. An analysis was carried out to compare the programs *H5P* (H5P, 2022) and *CBA ItemBuilder* (Kröhne 2023). In light of considerations around data protection, the availability of process data, and the promise of technical support from the Leibniz Institute for Research and Information in Education, the ECON 2022 authentic TBA was implemented using the CBA ItemBuilder.

ItemBuilder is an authoring tool for creating dynamic and interactive items for technology-based tests. The software was developed by the Centre for Technology-Based Assessment at the Leibniz Institute for Research and Information in Education. ItemBuilder allows editing of test items in a user interface and enables innovative item formats. Automatic scoring can be implemented using predefined rules. The delivery of final test environments from ItemBuilder can take place as a software package on a personal computer or USB stick, as a virtual machine, or online (Kröhne 2023). For the ECON 2022 project, assessment was delivered at schools using a USB stick. Automatic scoring took place after the assessment has been completed.

ItemBuilder was chosen because the availability of processing data would enable conclusions to be drawn about item processing. During processing, all user inputs and a time stamp are stored. One aim of the ECON 2022 assessment was to reconstruct test behaviour and the interaction between the test taker and the assessment. This is possible because computer-based test environments offer new possibilities for capturing and describing problem-solving processes due to the extensive data gained from the users' interactions with the test environment (Rausch et al. 2017, p. 569). Analyses of log data can make individual solution strategies visible (Rausch et al. 2017, p. 569). Log data is event based. Events are always linked to a test person and can refer to the content of the test or the test level (Kroehne and Goldhammer 2018, p. 533). User events such as the use of buttons, links, menu items, text input, or scrolling are made visible (Goldhammer et al. 2021). As the test person determines which interactions to carry out, it is possible to draw conclusions about their problem-solving strategies. Design and usability play an important role in log data analysis and influence the options for interpreting the log data (Kögler et al. 2020).

Assessment insight: incorporating authenticity regarding young people's economic opportunities for action

The initial focus of the ECON 2022 test development was the construction of authentic problem situations drawn from the lifeworld of the target group. The authentic test

situation was based on didactics of economics to ensure curricular validity (Fortunati and Winther 2023a). Based on the domain model developed for the project (Fortunati and Winther 2023a; Mislevy and Riconscente 2005), a framework for item development was created with the help of extensive curricular analyses which linked the content concepts to cognitive processes. Items were based on the content of the current curriculum for economics in NRW. Individual items refer to certain aspects of knowledge and capturing ways of processing. The economic competencies measured had to be found in the lifeworld and have clear relevance for managing everyday life. Display formats that would be authentic from the perspective of the lifeworld of 14-year-old students were researched in two ways. First, test instruments for measuring economic literacy were systematically researched and analysed; and second, relevant teaching and learning materials were researched to gain familiarity with the common display formats. In addition, they had to be anchored in the curriculum. When designing the items, care was taken to ensure that an immersive experience was possible. The term 'immersion' can be understood as the act of being completely absorbed into a virtual environment. Implementation using ItemBuilder enabled not only a realistic visualisation but also the inclusion of audio, video, and interactive elements to enhance authenticity. The possibility of experiencing digital content authentically in turn leads to high immersion (Wirth et al. 2007).

The modality of the ECON 2022 assessment included information intake, which took place visually and aurally. Depicting a situation close to the lifeworld was intended to enable the test persons to put themselves in the situation and identify with it. The action situations and work activities that were determined to be relevant for the target group were depicted in an authentic test environment. The didactic items involved a realistic representation of the test design tailored to the target group and a realistic representation of actual problems at item level. The digitalised design of the test enabled dynamic, innovative, and interactive item development. It was possible to embed multimedia content such as video and audio sequences (Jude and Wirth 2007, p. 49).

The ECON 2022 assessment was developed drawing on the characteristics of an authentic assessment as described by Gulikers et al. (2004) and Janesick (2006) (see "Principles of Authentic Assessments according to Janesick and Gulikers" section). The narrative of the ECON 2022 assessment depicts a concrete economic situation, namely a visit to a supermarket. Two protagonists, Kim (female) and Juri (male), are introduced to the target group as two school friends who are both 14 years old. Setting the protagonists' age at the age of the target group was designed to enable test participants to identify with the situation. In the test scenario, Kim and Juri are going grocery shopping in the supermarket. The framework situation of 'going shopping in the supermarket' is repeatedly interrupted, for example, by social media messages, associations to their schoolwork, or calls from class mates. These interruptions constitute eight individual units. A unit represents a coherent section of the test content and can consist of several items (Leutner et al. 2008). In the units, different economic problems are addressed in specific items (see Table 2). An item is the smallest element of analysis within the test (Leutner et al. 2008). The content of the domain model was implemented with specific items in each situation. Each unit contains two to six items. The 36 items represent the domain-related content and cognitive requirements in a balanced way.

Table 2 ECON 2022 Items

Units	Multiperspectivity	Items	No	Item content	Question type
Shopping list	Personal-financial, professional-entrepreneurial, socioeconomic	FT12	1	Needs and desires	Hotspot
		FT13	2	Econom. Difference organic products/ conv. Products	Multiple-choice
		FT14	3	Scarcity concept	Concept-mapping
Smartwatch	Personal-financial, professional-entrepreneurial	FT21	4	Digital marketing	Hotspot
		FT22	5	Impact of digital marketing strategies	Open response
		FT23	6	Impact of digital marketing strategies from an int. Perspective	Multiple-choice
Project work	Personal-financial, professional-entrepreneurial, socioeconomic	FT31	7	Sustainability definition	Open response
		FT32	8	Facets of sustainability	Drag and drop
		FT33	9	Fair trade products	Supplement
		FT34	10	Concept of fair trade	Matching
		FT35	11	Information sources about fair trade	Multiple-choice
Economic activities on a global scale	Personal-financial, professional-entrepreneurial	FT41	12	Annual interest calculation	Single-choice
		FT42	13	Compound interest	Single-choice
		FT43	14	Intra-year interest	Single-choice
		FT44	15	Tariffs and consequences for businesses	Multiple-choice
		FT45	16	Profit concept	Drag and drop
		FT46	17	Purchasing power	Multiple-choice
In-app purchase	Personal-financial	FT51	18	Contracts	Drag and drop
		FT52	19	Percentage calculation 1	Calculation
		FT53	20	Percentage calculation 2	Calculation
		FT54	21	Problems of in-app purchases	Open response
Bluetooth headphones	Personal-financial, professional-entrepreneurial	FT61	22	Subtraction calculation	Calculation
		FT62	23	Percentage calculation	Calculation
		FT63	24	Cost advantages of online shopping	Multiple-choice
		FT64	25	Terms of contracts	Drag and drop
		FT65	26	Currency conversion	Single-choice
		FT66	27	Online shopping	Multiple-choice
At the checkout	Personal-financial, socio-economic	FT71	28	Paying with change	Open response
		FT72	29	Functions of money	Single-choice
		FT73	30	Contract: yes, or no?	Single-choice
		FT74	31	Paying with a debit card	Open response
After the purchase	Socioeconomic	FT81	32	Pricing	Single-choice
		FT82	33	Economic cycle	Concept-mapping

Table 2 shows the eight economic situations (units), each representing distinct content emphases for the target audience. Each of the situations should be considered from multiple perspectives. In each unit, students are presented with multiple items that examine these situations from various perspectives. Alongside the item content, Table 2 also indicates the question type. Economic situations can be modelled using a linguistic-argumentative or mathematical-analytical approach. Economic literacy and numeracy are
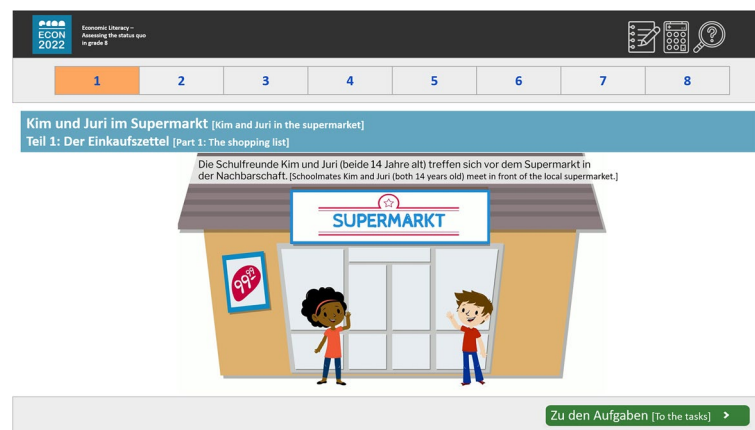
Fig. 2 Introduction to the authentic test environment

considered as domain-specific areas of economic knowledge that represent basic skills for (economic) vocational action (Winther 2010). The curricular representation of the domain model was examined by analysing 31 curricula for economic education at lower secondary school drawn from 10 different federal states and school types in Germany.

All situations of the items tie in with a video-based introduction to the framework situation and its progression between items. Sequencing makes it possible to structure complex issues individually (Bley et al. 2015). For example, in the ECON 2022 assessment, video sequences can be viewed repeatedly. Figure 2 shows a screenshot of the video, which contains spoken texts, sound, and subtitles (Jude and Wirth 2007; Finken et al. 2017). The test person emulates the typical stations that a visit to the supermarket entails.

In addition to developing an authentic framework situation, it was also important to develop stimulating items and to choose realistic item formats. Under the aspect of signalling—i.e. directing the focus to relevant elements—these were specifically highlighted. Unimportant details were omitted to avoid redundancy (Bley et al. 2015). It was crucial not only that the entire test situation could be found in the students' lifeworld, but also that individual items were authentic. Items were distinguished according to whether they represented an action situation that was directly drawn from the young people's everyday life, an action situation that would be accessible from the young people's perspective even if it was not directly drawn from their everyday life, or an action situation that was altogether unfamiliar or alien to the young people. In addition to visualising the structure of the assessment, Fig. 3 shows an example of the design of Item 1 from Unit 2. First, there is a video-based introduction to the item battery that constitutes Unit 2. The subjects are guided to the supermarket, where they receive a message on the imaginary social media platform Picturegram. An influencer, who is introduced as a favourite influencer, is promoting a smartwatch. The item was designed to explore young people's understanding of how online advertising strategies can influence purchasing decisions. Correct answer options are arguments that mention being directly addressed, belonging to a community, or pressure to act quickly because the offer is due to expire shortly. In curricular terms, the item can be assigned to content area 1, 'Economic activity in the market economy' (MSB 2019, p. 20). The aim of this area is to develop the
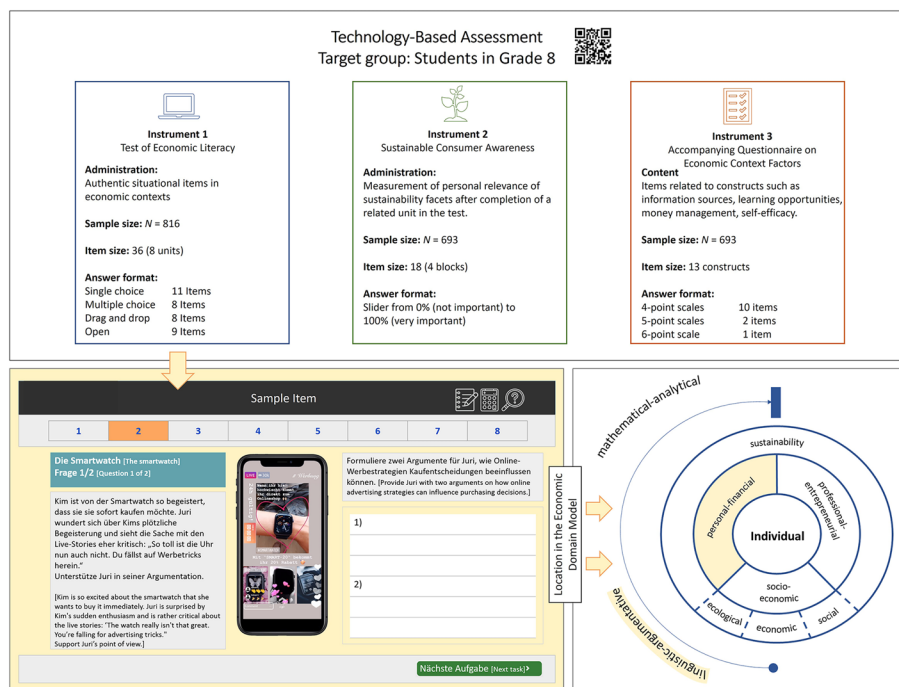
Fig. 3 Assessment overview and sample item: smartwatch

judgement competence of being able to assess the influence of advertising and social media on one's own consumer behaviour. In addition, the item can also be assigned to content area 8, 'Acting as consumers' (MSB 2019, p. 16); this content area deals with purchasing decisions in the digitalised world, whereby one focus is the influence of advertising on purchasing decisions. The item can be assigned to the personal-financial area in the domain model. Access is linguistic-argumentative. Scanning the QR code shown in Fig. 3 will give the reader access to the exemplary test environment.

In terms of difficulty-generating criteria, this item can be assigned to level 2 for the principles of specificity, cognition, and modelling. The criterion of specificity assesses the expertise that the item requires. The probability of solving this item is expected to be higher if students have attended lessons on economics-related subjects up to Grade 8. The criterion of cognition assesses the level of comprehension that the students have to demonstrate to answer the question correctly. To solve the item, individual solution steps must be applied. It is not possible to solve the item simply by reproducing pure factual knowledge. Modelling is also assigned to level 2 since the item contains audio-visual material that could distract from solving the item. For the criterion of authenticity, the item presented in Fig. 3 corresponds to level 1 and thus represents the students' lifeworld.

The targeted design of tasks can avoid cognitive overload (Sweller et al. 1998). In this context, continuity is based on the maximum processing capacity of humans and can be optimised by presenting task formats and contents in a systematic, clear, and well-structured manner (Bley et al. 2015). The ECON 2022 assessment is designed in a consistent, structured style and continues to offer helpful tools that can be easily opened. The assumption was that authentic situations and the simulation of familiar behaviours

would enable students to use their economic competencies more effectively in a TBA than in a paper-and-pencil questionnaire. Paper-and-pencil surveys require comprehensive descriptions to establish a connection with everyday reality. Too much reading can require excessive concentration and can promote cognitive overload and even failure at the items (Bley et al. 2015, p. 4). All materials were adapted to reduce the linguistic complexity and adjust them to the language competences of the target group. The materials were based on the lifeworld of the target group not only linguistically but also aesthetically (Bley et al. 2015). Establishing the lifeworld reference via videos instead of texts can lead to a reduction in the cognitive load of the subjects (Bley et al. 2015). Subtitles were used in the ECON 2022 assessment to engage equally those students with a high level of reading proficiency and those with reading difficulties. Students with reading difficulties could benefit because the audio-visual load was comparatively lower than the reading load for the same amount of information.

Regarding the psychometric quality of the test instrument, Fortunati and Winter (2023b) found sufficient empirical evidence of measurement accuracy for the construct of economic literacy. The following statements are based on field test data from the ECON 2022 assessment: the assessment can be evaluated as empirically reliable, valid, and fair for Grade 8 students. Adams and Khoo (1996) suggest a range of values between 0.75 and 1.33 for item fit (wMNSQ), while large-scale assessments like PISA consider stricter values between 0.85 and 1.15 as appropriate (OECD 2020). All items except FT22 meet the strict PISA value. The t values exhibit a dispersion ranging from -5.20 to 5.4. Furthermore, no significant gender or language differences were observed at test level in the differential item functioning (DIF) analyses. A significant DIF effect was found for both migration background and socioeconomic background. The DIF effect was 0.282 for migration background and 0.429 for socioeconomic background, both of which can be considered low according to the classification by Paek and Wilson (2011).

Validation and revision of the ECON 2022 assessment

Expert validation of the test

Content validation is an important part of test development, but it is often neglected (Ollesch et al. 2018, p. 129). The quality of a test depends to a large extend on the fulfilment of quality criteria. Specifically, the alignment of the theoretical construct with the actual test, in terms of validity, is important (Loerwald and Schnell 2016). To check the validity of the ECON 2022 test construction with a focus on both the authenticity of the assessment and item difficulties, the developed test items were validated using the framework of an expert rating. In order to select suitable candidates for the purpose of expert validation, researchers with research interests in test construction, economics, didactics, psychology, and competence development were invited to take the survey.

The resulting sample included a total of $N=25$ experts with expertise in test development ($n=10$), economics, i.e. economics or business education or business psychology ($n=11$), and schools and teaching ($n=12$). Individual experts could be assigned to two groups, indicating that they had attributed themselves expertise in two areas. The 25 experts thus represented expertise in the three fields of action. The validation study was based on the design criteria of specificity, cognition, modelling, and authenticity. Experts assessed the items using the associated decision trees.

The use of expert surveys allows validation of a model after development and implementation and thus to check the nature of decisions during test development (Offergeld 2011, p. 197). The validation study for the ECON 2022 project collected ratings based on the four design criteria (Beck 2020) and the usability of the assessment (Sangmeister et al. 2018). The expert ratings served as external verification of the authenticity of the ECON 2022 assessment and individual test items. Data analysis of the expert survey took place in two steps: first, the experts' rating of the four difficulty-generating criteria was analysed descriptively and quantitatively; and second, the expert rating was compared with a previously conducted self-rating. The analysis of free-text comments on individual items was carried out discursively and qualitatively by the test development team. Using both sources of information, the test development team reviewed the test items and fine-tuned their adaptation with the domain model. Table 3 lists the authenticity values arising from the expert rating.

Table 3 shows that the items were mostly assessed as authentic by the experts with the exception of items 8_1 and 8_2. Authenticity is not an all-or-nothing decision but a staged assessment in which we distinguish between basal and overall authenticity levels in a fluid transition. Therefore, we conducted a relative comparison of the items. Cut-off values were formed from the experts' ratings. According to the experts, only two items did not represent an action situation accessible to the students and were rated with a rather low level of authenticity. The values here were above 2.30. A further 16 items were given a rating of medium authenticity. The values here ranged from 1.51 to 2.30. The items therefore represent a situation that students could think of as potentially accessible for them in the future. Next, 17 items were rated as very authentic, i.e. as an action situation that reflected everyday life. Here the values were below 1.5. The two items which were rated as non-authentic by the experts required review and adaptation. However, since these were the last two items of the assessment and thus formed the conclusion of the test, it seemed justifiable to loosen their reference to the lifeworld even further to generate items that were more strongly geared towards reflection on economic systems. Selecting and adapting the items based on the expert ratings increased content validity.

Further, Table 3 reveals a difference in ratings between the expert and self-ratings in 18 of 34 items. It is striking that 16 of the 18 items that showed a deviation were rated as more authentic by the experts, and only two items were rated as less authentic in comparison to the self-rating. An example of an item that was rated as less authentic by the experts was item 1 from Unit 3. In this item, authenticity was generated by setting the protagonists, Kim and Juri, a homework item that involved designing a poster defining the term sustainability. In the self-rating, the item was assigned to level 1, indicating the item corresponded to an everyday action situation for the students. Overall, however, the expert rating for this item had a total value of $M = 1.76$, which would prompt the item to be assigned to level 2. The experts with economics expertise ($n = 11$) rated the item with $M = 1.55$. It can be assumed that these experts are more familiar with the students' lifeworld than the test developers ($n = 10$), who rated the item with $M = 2.2$. The experts with expertise in school and teaching ($n = 12$) rated the item with $M = 1.42$. This rating indicated that such items occur in the children's everyday school life. A limitation here was the fact that none of the experts had expertise solely of the young people's lifeworld. The development team interpreted the ratings results as a call for a revised

Table 3 Authenticity of the ECON 2022 assessment: expert ratings and self-ratings

Item	Expert rating (ER)				Self rating (ExpR)	Difference ER–ExpR	Difference		
	N	Min	Max	1 (high) to 3 (low) M	1 (high) to 3 (low) M	Total	Economic expertise	School and teaching	Test developement
1_1	24	1	3	1.21	1.00	− 0.08	0.90	0.82	0.70
1_2	25	1	3	1.48	2.00	− 1.46	− 0.45	− 0.42	− 0.40
1_3	24	1	3	1.50	2.00	− 0.08	0.40	0.55	0.60
1_4	23	1	3	1.87	2.00	0.35	1.11	1.10	1.10
2_1	23	1	1	1.00	2.00	− 0.13	1.00	1.00	1.00
2_2	25	1	2	1.44	1.04	− 0.36	0.55	0.67	0.40
2_3	25	1	3	1.60	3.00	− 0.96	− 0.27	− 0.33	− 0.90
3_1	25	1	3	1.76	1.00	− 0.40	− 0.55	− 0.42	− 1.20
3_2	24	1	3	2.13	2.00	− 1.09	− 0.30	0.09	− 0.20
3_3	25	1	2	1.36	2.00	− 0.24	− 0.36	− 0.25	− 0.50
3_4	25	1	3	2.08	3.00	− 0.52	0.00	0.25	− 0.50
3_5	25	1	3	1.88	2.00	− 0.16	− 0.82	− 0.75	− 1.00
4_1	25	1	3	2.16	2.00	− 0.04	0.00	− 0.08	− 0.30
4_2	25	1	3	2.16	2.00	− 0.48	− 0.09	− 0.08	− 0.20
4_3	25	1	3	2.12	2.00	− 0.12	− 0.09	− 0.25	− 0.10
4_4	23	1	3	2.30	3.00	0.46	0.70	0.70	0.60
4_5	25	1	3	2.04	2.00	− 0.42	− 1.18	− 1.00	− 0.90
4_6	25	1	3	2.12	3.00	− 0.71	− 0.18	− 0.08	0.00
5_1	24	1	3	1.29	2.00	− 0.88	− 0.20	− 0.17	− 0.33
5_2	25	1	3	1.16	3.00	0.08	0.91	1.00	0.70
5_3	25	1	3	1.24	3.00	0.08	0.82	1.00	0.80
5_4	25	1	2	1.16	1.00	− 0.32	0.82	0.83	0.80
6_1	25	1	1	1.00	1.00	0.04	1.00	1.00	1.00
6_2	24	1	2	1.25	3.00	0.08	0.82	0.83	0.70
6_3	25	1	3	1.64	2.00	− 1.20	0.55	0.58	0.10
6_4	24	1	2	1.13	1.00	0.08	1.00	0.82	0.90
6_5	25	1	3	1.68	1.00	− 0.17	0.45	0.17	0.50
6_6	25	1	3	1.52	2.00	− 0.21	− 0.64	− 0.67	− 0.40
7_1	25	1	3	1.36	3.00	0.48	0.36	0.67	0.40
7_2	24	1	3	1.71	3.00	− 0.64	− 0.70	− 0.64	− 0.90
7_3	24	1	3	1.38	2.00	− 0.16	0.80	0.82	0.50
7_4	23	1	2	1.04	1.00	− 0.40	0.00	− 0.10	0.00
7_5	25	1	3	1.28	2.00	− 0.16	0.73	0.92	0.40
8_1	25	1	3	2.48	3.00	− 0.68	− 0.64	− 0.50	− 0.30
8_2	25	2	3	2.56	3.00	0.36	0.36	0.33	0.60

definition of the item. In the item revision process, the team reviewed the item and considered scenarios of how a poster could be designed differently.

Item 5 from Unit 6 deals with currency conversion. A price comparison of headphones and the use of a currency calculator were modelled as an authentic situation. Here it is noteworthy that although the expert ratings can be assigned to level 2 overall, indicating that the action situation should be accessible to young people even though it may not be an everyday situation, the experts with expertise in school and teaching represented this opinion most strongly. The self-rating of this item generated an attribution to level 1

Table 4 Correlations between difficulty-generating characteristics

	Item difficulty (field test)	Specificity	Cognition	Modelling	Authenticity
Item difficulty (field test)	1				
Specificity	− .086	1			
Cognition	− .364*	.525**	1		
Modelling	− .176	.212	.365*	1	
Authenticity	− .269	.477**	.445**	.371*	1

*<.05, **<.01.

because it was assumed that with the rise of the internet and e-commerce and given the permanent use of social media and mobile devices, it is now easier than ever for people to shop online from retailers all over the world. This means that 14-year-olds might be interested in purchasing items from international retailers trading in a currency other than the students' homeland currency, and a currency calculator can help them understand how much items would cost in their local currency.

Table 3 also offers a comparison of ratings by experts' area of expertise. A closer look at the rating differences in Table 3 reveals a high level of agreement between the expert groups. The maximum difference between the three expert groups was above 0.5 for only four items. The greatest difference was between experts with expertise in school and teaching and experts in test development.

Authenticity as a Difficulty-Generating Characteristic

The realisation of an authentic test situation requires the implementation of multimedia content. Identifying with the setting should not generate difficulty for the test taker. This section examines whether authenticity is a difficulty-generating feature like specificity, cognition, and modelling, or whether it is purely a design criterion.

Table 4 shows the correlations between the expert ratings of the difficulty-generating characteristics and the expert rating of authenticity at item level ($N=34$). The results reveal that there was a medium strong correlation according to Spearman-Rho between authenticity and specificity (0.477**), cognition (0.445*), and modelling (0.371*), which was significant in all three cases. The low significance level can be explained by the limited number of only 25 expert ratings. Expert ratings were related to the difficulty of the individual test items. To analyse the data from the field test, a polytomous 1PL-IRT model, the multidimensional random coefficients multinomial logit model (Adams et al. 1997), was selected and scaled using ACER ConQuest (Adams et al. 2018).[1] According to theory, the three characteristics of specificity, cognition, and modelling should have showed a positive correlation with item difficulty; however, only cognition (-0.364*) showed a significant correlation with the measured item difficulty. As expected, authenticity (-0.269) showed no significant correlation with actual item difficulty.

---

[1] The supplementary methodology for data collection and data analysis of the ECON-Assessment can be found in the publication: Fortunati, F., Welsandt, N. J., Henicz, F., Abs, H. J., & Winther, E. (2024). Validierung des Testinstruments anhand der Feldtestdaten [Validation of the test instrument based on field test data] in Winther, E. & Abs, H. J. (Eds.), ECON 2022. Ökonomische Bildung in Jahrgang 8: Kompetenzen und Einstellungen [Economic Education in Grade 8: Design, Competencies, and Attitudes] (p.117–133). Published by Waxmann, CC BY 4.0. https://doi.org/10.31244/97838 30998389.4_3

Table 5 Correlations with dummy control variables

	Item difficulty (field test)
Item difficulty (field test)	1
Expert rating: Dummy variable authenticity	− .123
Self-rating: Dummy variable authenticity	.195
Total-rating: Dummy variable authenticity	.106

To exclude the possibility that the result was only an artefact based on rater bias, a second analysis was carried out with z-standardised expert assessments. Here, the mean value of all expert assessments in one criterion was set to 0, and the individual expert assessments were then included in standard deviation proportions. In this second analysis, authenticity was also independent of the measured difficulty (-0.41) and correlated with specificity (0.430*), cognition (0.403*), and modelling (0.342*).

To analyse the empirical correlation of authenticity with the measured item difficulty for the constructed test items, dummy variables were formed. The previously explained three levels of authenticity were compressed into two levels: *lifeworld relevance* and *no lifeworld relevance*. Levels 1 and 2 were combined and recorded as 1; level 3 was recorded as 2. Three dummy variables were formed with the mean values of the experts' ratings and the self-ratings, and the mean value of both and a correlation were calculated. The results also showed empirically that authenticity was independent of item difficulty; the finding was that authenticity was not perceived as a difficulty-generating feature (see Table 5).

Discussion and outlook

This article examined which processes have to be completed to create a TBA as an authentic assessment to construct a valid assessment for measuring economic literacy among students in the 8th grade in the federal state of NRW, Germany. The aim of the ECON 2022 assessment was to show which competencies the students already had at this stage and what competencies could therefore be expected from them at the beginning of their training in a vocational area. The added value that the use of technology can have for innovative and interactive item development was highlighted.

Relevance is evident from the fact that TBA has repeatedly been used in accordance with the possibilities it offers when *recording* economic literacy. New possibilities for *measuring* knowledge, skills, and abilities, for example through the use of innovative response formats, have not yet been sufficiently exploited. In constructing the ECON 2022 assessment, care was taken to ensure that the test environment had curricular validity in that it was aligned with the curriculum of the subject of economics in the state of NRW (Fortunati and Winther 2023a). Test design was based on the basic parameters of the authentic assessment and on the domain model, which is based on a model of evidence-centred design. An analysis of the ECON 2022 assessment illustrated that an orientation towards the difficulty-generating criteria of specificity, cognition, and modelling, combined with authenticity and usability in the construction of a TBA, leads to a valid assessment.

For economic literacy, TBAs provide the opportunity realistically to construct typical work and thought processes. In constructing new test instruments, the focus lies on implementing an environment that is as authentic as possible and thus significant for personal learning and the living environment, which can have emotional and motivational effects through the use of media.

The study shows that authentic simulations can depict processes and actions that illustrate the everyday life of students and thus provide an accurate insight into students' knowledge and skills at the beginning of vocational training in economics. Using the simulated scenario of grocery shopping with the guidance through the assessment by the protagonists Kim and Juri, the action- and comprehension-based ability structures of the students who completed the test could be recorded. Moreover, the simulation established a reference to real-life action processes in the field of economics.

One limitation of the analysis of the ECON 2022 test design is that the target group of students was not included in the rating of authenticity. A further limitation is that no immersion was incorporated in the assessment, i.e. the participants themselves cannot completely merge with the test situation. Immersion would be conceivable in a virtual reality environment. Moreover, in an augmented reality environment, it would also be possible to ask questions in a virtual supermarket that has been specifically prepared for this purpose. That said, the aim of the ECON 2022 test design was not to reproduce reality through an immersive experience, but to construct an assessment that was aligned with the experienced lifeworld of the test persons and to model items that fit the theoretical construct. In addition, the aim was to map the current state of economic literacy into a large-scale assessment. In this context, the use of virtual or augmented reality scenarios was deemed too costly for the purpose.

The expert survey showed no significant correlation between authenticity and the item difficulties measured in the field test. Authenticity was objectively not a difficulty-generating characteristic. This result should be interpreted positively as authenticity was not supposed to have a difficulty-generating effect. This finding, in turn, offers proof of the quality of the implementation of the authentic test environment. After the results of the expert validation have been incorporated into the final adaptation of the test items for the ECON 2022 assessment, the main study will seek to measure the actual status *before* vocational training so that teachers can address what commercial competencies the students already have when they enter vocational training and what can be expected of the students in this context at the beginning of vocational training.

The ECON 2022 assessment can claim to have implemented a precise understanding of competence in economic literacy in authentic situations, and to have mapped various facets of economic competence.

This study aimed to develop and validate a technology-based authentic assessment that can be used as a theoretical basis for measuring economic competencies prior to entry into vocational training occupations. For vocational education and training, economic literacy acts as a condition for the development of area-specific competencies for trainees. In NRW, the vocational preparation of students is becoming more central and has also been a curricular component since 2020/2021 in the form of the subject of economics. The study focuses on the new curriculum component, which should better prepare students for the profession and provide connections with commercial apprenticeships.

The newly developed authentic TBA—ECON 2022—enables to assess economic literacy in schools in a way that maintains curricular validity, and to analyse what can be expected of students when they enter a vocational training occupation.

Declarations

Competing interests
The authors declare that they have no competing interests.

References
Achtenhagen F, Winther E (2008) Wirtschaftspädagogische Forschung zur beruflichen Kompetenzentwicklung [Research in economics education on developing vocational competence]. In Bundesministerium für Bildung und Forschung [Federal Ministry of Education and Research] (Ed.), *Kompetenzerfassung in pädagogischen Handlungsfeldern: Theorien, Konzepte und* Methoden. Bundesministerium für Bildung und Forschung, pp 117–140

Achtenhagen F, Winther E (2009) Konstruktvalidität von Simulationsaufgaben: Computergestützte Messung berufsfachlicher Kompetenz—am Beispiel der Ausbildung von Industriekaufleuten [Construct validity of simulation tasks]. Bericht an das Bundesministerium für Bildung und Forschung. *Professur für Wirtschaftspädagogik der Georg-August Universität Göttingen*. Göttingen.

Ackermann N (2018) Dokumentation des revidierten Tests zur Wirtschaftsbürgerlichen Kompetenz (WBK-T2): Item-Spezifikation, Item-Kennwerte, Koderungsmanual [Documentation of the revised Test of Economic Civic Competence (WBK-T2): Item specification, item characteristics, coding manual] [Unpublished manuscript]. Zurich University of Teacher Education.

Ackermann N (2019) *Wirtschaftsbürgerliche Kompetenz Deutschschweizer Gymnasiastinnen und Gymnasiasten: Kompetenzmodellierung, Testentwicklung und evidenzbasierte Validierung [The economic civic competence of Swiss-German secondary school pupils]*. [Doctoral dissertation]. University of Zurich. https://doi.org/10.5167/uzh-175377

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. AERA, APA, NCME.

Aziz M, Yusoff N, Yaakob M (2020) Challenges in using authentic assessment in 21st century ESL classrooms. Int J Eval Res Educ 9(3):759–768

Beck K (1989) 'Ökonomische Bildung'—Zur Anatomie eines wirtschaftspädagogischen Begriffs ['Economic education']. Zeitschrift Für Berufs- Und Wirtschaftspädagogik 85(7):581

Beck K (2020) Ensuring content validity of psychological and educational tests—the role of experts. Front Learn Res 8(6):1–37. https://doi.org/10.14786/flr.v8i5.517

Beck K, Krumm V (1998) Wirtschaftskundlicher Bildungstest (WBT). Handanweisung [Economic literacy test]. Hogrefe Verlag für Psychologie.

Beck U, Brater M, Tramsen E (1976) Beruf, Herrschaft und Identität. Ein subjektbezogener Ansatz zum Verhältnis von Bildung und Produktion [Occupation, domination, and identity]. Soziale Welt 4(1):8–44

Birindiba Batista I, Hahn-Laudenberg K, Abs HJ (2022) *Bildung für nachhaltige Entwicklung im Schulcurriculum [Education for sustainable development in the school curriculum]. Eine fächer- und schulformübergreifende Analyse.* [Manuscript in preparation]. Faculty of Pedagogy, University of Leipzig.

Bley S, Wiethe-Körprich M, Weber S (2015) Formen kognitiver Belastung bei der Bewältigung technologiebasierter authentischer Testaufgaben – eine Validierungsstudie zur Abbildung von beruflicher Kompetenz [Forms of cognitive load when completing technology-based authentic test items]. ZBW 111(2):268–294

Bloom B, Englehart M, Furst E, Hill W, Krathwohl D (1956) Taxonomy of educational objectives: handbook i: the cognitive domain. Longman, UK

Brandlmaier E, Frank H, Korunka C, Plessnig A, Schopf C, Tamegger K (2006) Ökonomische Bildung von Schüler/innen Allgemeinbildender Höherer Schulen [Economic literacy of secondary school students]. Facultas.

Welsandt *et al. Empirical Res Voc Ed Train*   (2024) 16:4

Page 25 of 27

Breitbach E, Wagner J (2018) Family matters: Financial literacy and the incoming college freshman. In: M Förster, R Happ, WB Walstad, CJ Asarta, (Eds), Financial Literacy (Empirische Pädagogik, 32(3/4), Themenheft). Verlag Empirische Pädagogik.

Deutsches Institut für Erwachsenenbildung (2015). Stimmt's-Kärtchen. Rechnen im Bereich Finanzielle Grundbildung. ['Did you get it right?' flash cards. Maths in the area of basic financial education].

Corsten H, Roth S (2012) Nachhaltigkeit als integriertes Konzept [Sustainability as an integrated concept]. In: H Corsten, S Roth, *Nachhaltigkeit*. Gabler Verlag. pp 1–13, https://doi.org/10.1007/978-3-8349-3746-9_1

Ersozlu Z, Ledger S, Hobbs L (2021). Virtual simulation in 1TE. Technology driven authentic assessment and moderation of practice. In: T Barkatsas, T McLaughlin (Eds), Authentic assessment and evaluation approaches and practices in a digital era: A kaleidoscope of perspectives. Brill, pp. 53–68. https://doi.org/10.1163/9789004501577

Finken J, Marx F, Meyer M, Krieter P, Breiter A (2017) Entwicklung und Durchführung computerbasierter Tests zur Messung von Musikkompetenzen [Developing and implementing computer-based tests for measuring competence in music]. In: Igel C et al. (Eds.), Bildungsräume, DeLFI 2017 – Die 15. E-Learning Fachtagung Informatik/ Lecture in Informatics (LNI). Gesellschaft für Informatik, pp 63–74

FINRA Investor Education Foundation (2018) 2018 National Financial Capability Study State-by-state Survey instrument. https://finrafoundation.org/sites/finrafoundation/files/NFCS-2018-State-by-State-Methodology.pdf

Fortunati F, Winther E (2023a) Intensionen und Intentionen von Curricula: Domänenmodelle als Voraussetzungen für die Kohärenz instruktionaler Aktivität in geringstrukturierten Domänen am Beispiel der ökonomsichen Bildung [Intensions and intentions of curricula: Domain models as conditions for the coherence of instructional activity in low-structure domains using the example of economics education]. [Manuscript in preparation, accepted, 20.03.2023]. Zeitschrift für Unterrichtswissenschaft.

Fortunati F, Winther E (2023b) Curriculare Analysen als Baustein der Assessmentkonstruktion [Curricular analyses as a component of assessment construction]. [Manuscript in preparation, accepted, 09.02.2023]. *Zeitschrift für Erziehungswissenschaft*.

Fortunati F, Welsandt NJ, Abs HJ, Winther E (2024) Die Entwicklung eines authentischen technologiebasierten Tests zur Erfassung wirtschaftlicher Kompetenz. Psychometrische Eigenschaften des Testinstruments TBA-EL [The development of an authentic technology-based test to assess economic literacy. Psychometric characteristics of the TBA-EL test instrument]. Zeitschrift für Berufs- und Wirtschaftspädagogik.

Gelman R, Greeno JG (1989) On the nature of competence: Principles for understanding in a domain. In: Resnick LB (Ed.), *Knowing, learning and instruction. Essays in honor of Robert Glaser*. Lawrence Erlbaum Associates, pp 125–186

Goldhammer F, Kroehne U (2020) Computerbasiertes Assessment [Computer-based assessment]. In: H. Moosbrugger & A. Kelava (Eds.), Testtheorie und Fragebogenkonstruktion [Test theory and questionnaire construction]. Springer. [Test theory and questionnaire construction]. Springer. pp 119–141. https://doi.org/10.1007/978-3-662-61532-4

Goldhammer F, Scherer R, Greiff S (2020) Advancements in technology-based assessment: Emerging item formats, test designs, and data sources. Front Res Topics. https://doi.org/10.3389/978-2-88963-506-1

Goldhammer F, Hahnel C, Kroehne U (2020) Analysing log file data from PIAAC. In: Maehler D, Rammstedt B (Eds.), *Large-scale cognitive assessment. Methodology of educational measurement and assessment*. Springer, pp 239–269. https://doi.org/10.1007/978-3-030-47515-4_10

Goldhammer F, Hahnel C, Kroehne U, Zehner F (2021) From byproduct to design factor: on validating the interpretation of process indicators based on log data. Large-Scale Assess Educ. https://doi.org/10.1186/s40536-021-00113-5

Gulikers J, Bastiaens T, Kirschner P (2004) A five-dimensional framework for authentic assessment. Education Tech Research Dev 52(3):67–85

de Haan G, Holst J, Singer-Brodowski M (2021) Berufliche Bildung für nachhaltige Entwicklung: Genese, Entwicklungsstand und mögliche Transformationspfade [Vocational education for sustainable development]. BWP 50(3):10–14

Hartig J, Rauch D (2008) Psychometric models for the assessment of competencies. In: Hartig J, Klieme E, Leutner D (eds) Assessment of competencies in educational contexts. Hogrefe, pp 69–90

Hartig J, Roczen N (2020) SysKo. SysKo-BNE—Messung von Systemkompetenz als Indikator im Bereich Bildung für nachhaltige Entwicklung [Measurement of systems competence as an indicator of sustainable educational development]. [Unpublished test material]. Leibnitz Institut für Bildungsforschung und Bildungsmethoden .

von Hering R, Zingelmann H, Heinze A, Lindmeier A (2020) Lerngelegenheiten mit kaufmännischem Kontext im Mathematikunterricht der allgemeinbildenden Schule—Eine Schulbuch- und Aufgabenanalyse [Learning opportunities with a commercial context in mathematics lessons in general education schools—a textbook and task analysis]. ZE [j Educ Sci] 23(1):193–213. https://doi.org/10.1007/s11618-019-00925-w

von Hering R, Rietenberg A, Heinze A, Lindmeier A (2021) Nutzen Auszubildende bei der Bearbeitung berufsfeldbezogener Mathematikaufgaben ihr Wissen aus der Schule? Eine qualitative Untersuchung mit angehenden Industriekaufleuten [Do trainees use their knowledge from school when working on math problems related to their occupational field? A qualitative study with prospective industrial clerks]. JMD 42(2):459–490. https://doi.org/10.1007/s13138-021-00181-8

H5P (2022) Create, share and reuse interactive HTML5 content in your browser. https://h5p.org/.

Janesick VJ (2006) Authentic assessment primer. Peter Lang Publishing, Lausanne

Jude N, Wirth J (2007) Neue Chancen bei der technologiebasierten Erfassung von Kompetenzen [New opportunities in technology-based measuring of competences]. In: Hartig J, Klieme E (eds) Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik. Bundesministerium für Bildung und Forschung, pp 49–56

Kaiser T, Oberrauch L, Seeber G (2019) Measuring economic competence of secondary school students in Germany. ZBW – Leibniz Information Centre for Economics.

Klieme E, Hartig J, Rauch D (2008) The concept of competence in educational contexts. In: Hartig J, Klieme E, Leutner D (eds) Assessment of competencies in educational contexts. Hogrefe, pp 3–22

Klotz V (2015) Diagnostik beruflicher Kompetenzentwicklung. Eine wirtschaftsdidaktische Modellierung für die kaufmännische Domäne [Diagnostic for the development of professional competence]. *Springer Gabler.*

Klotz VK, Winther E, Festner D (2015) Modellierung der Entwicklung beruflicher Kompetenz: ein psychometrisches Modell für Wirtschaftsdomänen [Modelling the development of professional competence]. Berufungen Und Lernen 8(3):247–268

KMK. [Standing Conference of Ministers of Education and Cultural Affairs] (2016). Orientierungsrahmen für den Lernbereich Globale Entwicklung im Rahmen einer Bildung für nachhaltige Entwicklung [Framework for the learning area of global development in the context of education for sustainable development]. *Ständige Konferenz der Kultusminister der Länder*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2015/2015_06_00-Orientierungsrahmen-Globale-Entwicklung.pdf.

Knoll MAZ, Houts CR (2012) The financial knowledge scale: an application of item response theory to the assessment of financial literacy. J Consumer Affairs 36(3):381–410. https://doi.org/10.1111/j.1745-6606.2012.01241.x

Koh K (2017) Authentic assessment. Oxford Res Encycl Educ. https://doi.org/10.1093/acrefore/9780190264093.013.22

Kögler K, Rausch A, Niegemann H (2020) Interpretierbarkeit von Logdaten in computerbasierten Kompetenztests mit großen Handlungsräumen [Scope for interpreting log data in computer-based competency assessments with wide action remits]. https://www.bwpat.de/profil6_wuttke/koegler_etal_profil6.pdf.

Kröhne, U. (2023). Open Computer-based Assessment with the CBA ItemBuilder. https://cba.itembuilder.de

Kroehne U, Goldhammer F (2018) How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. Behaviormetrika 45(2):527–563. https://doi.org/10.1007/s41237-018-0063-y

Leahy W, Sweller J (2011) Cognitive load theory, modality of presentation and the transient information effect. Appl Cogn Psychol 25:943–951. https://doi.org/10.1002/acp.1787

Lempert W (2009) Berufliche Sozialisation. Persönlichkeitsentwicklung in der betrieblichen Ausbildung und Arbeit [Professional socialisation]. Schneider.

Leutner D, Hartig J, Jude N (2008) Measuring competencies: Introduction to concepts and questions of assessment in education. In: Hartig J, Klieme E, Leutner D (eds) Assessment of competencies in educational contexts. Hogrefe, pp 177–192

Loerwald D, Schnell C (2016) Diagnostik im Dilemma zwischen fachdidaktischen Ansprüchen und empirischen Anforderungen. Zur (vermeintlichen) Trivialität von Testitems [Diagnostics in the dilemma between didactic demands and empirical requirements]. In: zeitschrift für didaktik der gesellschaftswissenschaften (1), S. 58–73.

Mandell L (2008) The financial literacy of young American adults. Results of the 2008 National Jump$tart Coalition. Survey of high school seniors and college students. https://views.smgww.org/assets/pdf/2008%20JumpStart%20Financial%20Literacy%20Survey.pdf

Marzano RJ, Kendall JS (2007) *Die neue Taxonomie von Bildungszielen [The new taxonomy of educational goals]*. Corwin Press.

May H (Ed.) (2008) *Handbuch zur Ökonomischen Bildung [Handbook for economics education]*. De Gruyter Oldenbourg (9th ed.).

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2019) *Kernlehrplan fürdie Sekundarstufe I des Gymnasiums in Nordrhein-Westfalen. Wirtschafts-Politik [Core curriculum for the lower key stage of secondary education in North Rhine-Westphalia]*. https://www.schulentwicklung.nrw.de/lehrplaene/lehrplan/215/g9_wipo_klp_%203429_2019_06_23.pdf.

Ministerium für Arbeit, Gesundheit und Soziales des Landes Nordrhein-Westfalen (2020) *Kein Abschluss ohne Anschluss. Übergang Schule – Beruf in NRW. Handbuch zur Umsetzung der Standardelemente und Angebote [Guaranteeing next steps for school leavers]*. https://www.berufsorientierung-nrw.de/wp-content/uploads/2022/01/handbuch-kaoa-final-2020.pdf.

Ministerium für Schule und Bildung des Landes Nordrhein-Westfalen (2021) *Schulfach Wirtschaft [School subject: Economics]*. https://www.schulministerium.nrw/schule-bildung/schulpolitik/schulfach-wirtschaft.

Mislevy RJ, Riconscente MM (2005) *Evidence-centered assessment design: Layers, structures, and terminology*. SRI International.

Mudzingiri C (2019) The impact of financial literacy on risk and time preferences and financial behavioural intentions. [Doctoral dissertation]. University of the Free State. https://scholar.ufs.ac.za/xmlui/handle/11660/10268

National Center for Education Statistics (2013) The nation's report card: Economics 2012 (NAEP). Institute of Education Sciences, U.S. Department of Education.

Nicolini G (2012) Financial education online: Does it work? Facultà di Economia, Università degli studi di Roma.

Offergeld T (2011) *Wirtschaftlichkeit von Immobilien im Lebenszyklus [The economic efficieny of real estate in the lifce cyle.]* Gabler Verlag & Springer Fachmedien.

Ollesch J, Dörfler T, Vogel M (2018) Die inhaltliche Validierung von Unterrichtsvignetten durch eine mehrstufige Expertenbefragung [Content validation of lesson vignettes through multi-stage expert assessment]. In: Rutsch J, Rehm M, Vogel M, Seidenfuß M, Dörfler T (eds) Effektive Kompetenzdiagnose in der Lehrerbildung. Springer, pp 129–151

Organisation for Economic Co-operation and Development (OECD) (2014) PISA 2012 results: Students and money: Financial literacy skills for the 21st century (Volume VI). OECD Publishing.

Organisation for Economic Co-operation and Development (OECD) (2019). PISA 2021 Financial Literacy Analytical and Assessment Framework. https://www.oecd.org/pisa/sitedocument/PISA-2021-Financial-Literacy-Framework.pdf.

Organisation for Economic Co-operation and Development (OECD) (2020) PISA 2018 results: are students smart about money? (Volume IV). OECD Publishing.

Pellegrino JW (2012) The design of an assessment system focused on student achievement. A learning sciences perspective on issues of competence, growth, and measurement. In: Bernholt S, Neumann K, Nentwig P (Eds.), *Making it tangible—Learning outcomes in science education*. Waxmann, pp. 79–107

Pellegrino JW, Chudowsky N, Glaser R (eds) (2001) Knowing what students know—The science and design of educational assessment. National Academy Press, Washington

Rausch A, Kögler K, Frötschl C, Bergrab M, Brandt D (2017) Problemlöseprozesse sichtbar machen: analyse von Logdaten aus einer computerbasierten Bürosimulation [Making problem-solving processes visible]. ZBW 113(4):569–594

Welsandt *et al. Empirical Res Voc Ed Train* (2024) 16:4

Page 27 of 27

Rebmann K, Schlömer T (2020) Berufsbildung für eine nachhaltige Entwicklung [Vocational training for sustainable development]. In R. Arnold, *Handbuch Berufsbildung [Handbook vocational training]*. Springer VS, pp 325–337

Sangmeister J, Winther E, Deutscher V, Bley S, Kreuzer C, Weber S (2018) Designing Competence assessment in VET for a digital future. In: Ifenthaler D (Ed.), *Digital workspace learning*). Springer International Publishing. pp 55–92. https://doi.org/10.1007/978-3-319-46215-8_5

Sayers A (2007) Tips and tricks in performing a systematic review: reference management and identifying search terms and keywords. Br J Gen Pract 58(547):136. https://doi.org/10.3399/bjgp08X277168

Schank C, Lorch A (2018) Der Nachhaltigkeitsbürger in der soziökonomischen Bildung. Überlegungen zu einem wirtschaftsethisch fundierten soziökonomischen Bildungsideal soziökonomischen Bildungsideal [The sustainable citizen in socioeconomic education]. In: Engartner T, Fridrich C, Graupe S, Hedtke R, Tafner G (Eds.), Soziökonomische Bildung und Wissenschaft. Soziökonomische Bildung und Wissenschaft. Springer Fachmedien, pp 215–242

Schumann S, Eberle F (2014) Ökonomische Kompetenzen von Maturandinnen und Maturanden (OEKOMA) [Economic competencies of students leaving secondary school]. [Unpublished manuscript]. Department of Economics/Institute of Education, University Konstanz/Zürich.

Schroeder NL, Nesbit JC, Anguiano CJ, Adesope OO (2018) Studying and constructing concept maps: a meta-analysis. Educ Psychol 30:431–455. https://doi.org/10.1007/s10648-017-9403-9

Seeber S, Schumann S, Nickolaus R (2014) Ökonomische Kompetenzen: Konzeptuelle Grundlagen und empirische Befunde [Economic competences]. In: Weisseno G, Schelle C, (Eds.), *Empirische Forschung in gesellschaftswissenschaftlichen Fachdidaktiken – Ergebnisse und Perspektiven*. Springer VS, Wiesbaden. pp. 169–184. https://doi.org/10.1007/978-3-658-06191-3_12.

Steger D (2019) Technology-based assessment: A theoretical framework, psychometric modeling, and substantial issues in the assessment of cognitive abilities. [Doctoral dissertation]. Human Sciences and Education, Otto-Friedrich-University. https://doi.org/10.20378/irb-47740

Sweller J, van Merriënboer J, Paas F (1998) Cognitive architecture and instructional design. Educ Psychol Rev 10(3):251–296

Tanner S (2001) Authentic assessment: a solution, or part of the problem? High School J 85(1):24–29. https://doi.org/10.1353/hsj.2001.0020

Tóth M, Lančarič D, Savov R (2015) Impact of education on the financial literacy. [Unpublished manuscript]. Faculty of Economics and Management, Slovak University of Agriculture in Nitra.

Villarroel V, Bloxham S, Bruna D, Bruna C, Herrera-Seda C (2018) Authentic assessment: creating a blueprint for course design. Assess Eval High Educ 43(5):840–854. https://doi.org/10.1080/02602938.2017.1412396

Walstad WB, Robson D (1990) Basic economic test: Examiner's manual, 2nd edn. Joint Council on Economic Education, New York

Walstad WB, Soper JC (1998) Test of economic knowledge Examiner's manual. National Council on Economic Education, New York

Walstad WB, Rebeck K (2005) Financial fitness for life Upper elementary test. Examiner's manual. National Council on Economic Education, New York

Walstad WB, Rebeck K (2016) Test of financial literacy. Examiner's manual. Council for Economic Education, New York

Walstad WB, Watts M, Rebeck K (2006) Test of understanding in college economics: Examiner's manual, 4th edn. National Council on Economic Education, New York

Walstad WB, Rebeck K, Butters RB (2013) Test of economic literacy. J Econ Educ 44(3):298–309

Welsandt NJ, Abs HJ (2023) Testing economic literacy: an overview of measurement instruments of the past 30 years. J Soc Sci Educ. https://doi.org/10.11576/jsse-5855

Winther E (2010) Kompetenzmessung in der beruflichen Bildung [Measuring competence in vocational education]. W. Bertelsmann Verlag. https://doi.org/10.3278/6004148w

Winther E, Achtenhagen F (2008) Kompetenzstrukturmodell für die kaufmännische Bildung. Adaptierbare Forschungslinien und theoretische Ausgestaltung [A model of competence structure for business education]. ZBW 104(4):511–538

Winther E, Achtenhagen F (2009) A contribution to an international large-scale assessment on vocational education and training. Empir Res Voca Educ Train 1:85–102. https://doi.org/10.25656/01:5159

Winther E, Klotz V (2015) CERAFORMA test material. [Unpublished manuscript]. Department of Vocational Education and Training, University of Duisburg-Essen.

Winther E, Seeber S, Weber S, Bley S, Festner D, Kreuzer C, Rudeloff M, Sangmeister J, Wiethe-Körpich M (2016) ALUSIM test material. [Unpublished manuscript]. Department of Vocational Education and Training, University of Duisburg-Essen.

Winther E, Paeßens J, Tröster M, Bowien-Jansen B (2022) Immersives Lernen für Geringliteralisierte [Immersive learning for individuals with low levels of literacy] Chancen der Augmented Reality am Beispiel der Finanziellen Grundbildung. MedienPädagogik. 47:267–287. https://doi.org/10.21240/mpaed/47/2022.04.13.X

Wirth W, Hartmann T, Böcking S, Vorderer P, Klimmt C, Schramm H (2007) A process model of the formation of spatial presence experiences. Media Psychol 9(3):493–525. https://doi.org/10.1080/1521326070128307

Wobker I, Lehmann-Waffenschmidt M, Kenning P, Gigerenzer G (2012) What do people know about the economy? A test of minimal economic knowledge in Germany. Dresden Discussion Paper Series in Economics, No. 03/12, Technische Universität Dresden, Fakultät Wirtschaftswissenschaften, Dresden.

Publisher's Note

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