Historical Thinking Culture and Education

DOI 10.12685/htce.2222

Towards psychometrically sound assessments of knowledge, empathy and competencies in antisemitism-critical education

Progress and challenges

Lisa Zachrich [*a], *Jakob Baier*[b], *Volker Beck* [c], *Deidre Berger*[c], *Nicola Brauch* [d], *Jana-Andrea Frommer*[e], *Marc Grimm* [f], *Jana Habig*[d], *Daliah Zora Hauch*[d], *Sarah Jadwiga Jahn* [g], *Jonna-Margarethe Mäder* [d], *Marcel Mierwald* [h], *Katharina Netzer* [i], *Sarah Vosding*[d], *Ulrich Trautwein* [a]

[a]*University of Tübingen, Tübingen;* [b]*University of Bielefeld, Bielefeld;* [c]*Tikvah Institute gUG, Berlin;* [d]*Ruhr-University Bochum, Bochum;* [e]*Thuringian Police Training Center, Meiningen;* [f]*University of Wuppertal, Wuppertal;* [g]*University of Applied Sciences for Police and Public Administration in North Rhine-Westphalia, Dortmund;* [h]*Georg-Eckert-Institute, Braunschweig;* [i]*University of Mannheim, Mannheim;*

Abstract

Rising antisemitism threatens cohesion and peace in democratic societies. Teachers are expected to uphold democratic values as part of their professional responsibility and thus play a crucial role in preventing and addressing antisemitism. Yet, little is known about the effectiveness of educational interventions that aim to foster knowledge, competencies, and other potentially relevant constructs for antisemitism-critical education, such as empathy. This gap partly stems from a lack of adequate measurement instruments. This article describes the development and validation of a psychometrically sound instrument to measure these constructs in prospective teachers. In three steps, we report on (1) item development and initial testing; (2) a study with university students (N = 526); and (3) a validation with prospective teachers (N = 3,721[1]). Our results document the feasibility of constructing a knowledge test with acceptable psychometric properties and significant progress in measuring empathy, whereas more work is needed on assessing competencies.

1 The comma is used as a thousands separator.

* **Contact:** Lisa Zachrich lisa.zachrich@uni-tuebingen.de

University of Tübingen, Postdoc at the Hector Research Institute of Education Sciences and Psychology, Tübingen, Germany

© 2026 The Author(s). This is an Open Access article distributed under the terms of the Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0, http://creativecommons.org/licenses/by-sa/4.0/).

Keywords

Test development, antisemitism-critical education, knowledge test, empathy, antisemitism

1 Introduction

A rising tide of antisemitism in democratic societies threatens the security of Jewish people, democratic cohesion, and social peace (e.g., Bar-Halpern & Wolfman, 2025; Chernivsky & Lorenz-Sinai, 2026). Antisemitism is "a certain perception of Jews, which may be expressed as hatred toward Jews" (International Holocaust Remembrance Alliance [IHRA], 2016, p. 1). It may be manifested in actions and words against people, their belongings, locations, and institutions associated with Judaism and Jewish culture, as well as the state of Israel. Antisemitism poses a serious threat, requiring appropriate responses from all members of society (Bernstein et al., 2022). Schools deserve particular attention, as societal phenomena often unfold "as if under a magnifying glass" (Bernstein, 2020, p. 15; translated by authors) in schools. Teachers face increasing instances of antisemitism among students, ranging from implicit biases to expressions of national extremism (Chernivsky, 2025). Teachers in general and history teachers specifically are required to prevent and address antisemitism effectively (e.g., Tribukait, 2021). It is therefore essential to establish effective programs to support antisemitism-critical education (Chernivsky & Lorenz-Sinai, 2021, 2026).

Empirical studies show that many teachers feel insecure about addressing antisemitism in schools and not well prepared to address it after finishing their teacher training program. That could lead to downplaying critical incidents related to antisemitism, attributing them to developmental factors such as "a symptom of puberty" (Chernivsky & Lorenz-Sinai, 2021, p. 112), and focusing more on perpetrators than victims. Bernstein (2020) found that teachers usually avoid confronting such incidents, preferring to overlook them rather than discuss such topics they find uncertain or unfamiliar. This reluctance potentially stems from limited knowledge about antisemitism and its historical, social, and pedagogical contexts, leaving teachers unprepared to respond effectively (Bernstein, 2020; Chernivsky, 2025; Chernivsky & Lorenz-Sinai, 2026). For history teachers particularly, studies show that teachers themselves have misconceptions about the Shoah (e.g., Tribukait, 2021), and Holocaust Education has not fulfilled expectations with regard to combating antisemitism (e.g., Niedick et al., 2026; Pistone et al., 2024). In response, teachers, scholars, and policymakers have called for integrating antisemitism-critical education into teacher training, expanding professional development, and creating practical, evidence-based guidelines (Striewski, 2026; Thein & Hiepel, 2024; Vernikoff et al., 2024; Zentralrat der Juden in Deutschland et al., 2025).

Numerous educational interventions and teaching materials have been developed to support teachers in building confidence, knowledge and competence in addressing antisemitism. Despite substantial investments in such initiatives, their effectiveness remains uncertain (Bauer, 2021; Niedick et al., 2026). One key reason is the scarcity of validated measurement instruments—tools essential for rigorous effectiveness evaluations (Niedick et al., 2026). In response, the present study aims to: (1) identify current progress and persistent challenges in developing standardized assessment instruments for antisemitism-critical education, (2) develop a standardized measurement tool to assess prospective teachers' knowledge, empathy, and competencies for the prevention and intervention of antisemitism; and (3) present preliminary versions of test items for use in empirical research and future refinement.

2 Theoretical background

2.1 Requirements for antisemitism-critical education: Knowns and unknowns

What is needed in schools and education more broadly to successfully combat antisemitism? Research on teaching sensitive historical issues may help to better understand how to engage in antisemitism-critical education. Sensitive historical issues are topics that are emotionally

charged, politically contested, and closely connected to questions of identity. Such topics include, among others, colonialism, slavery, migration, National Socialism, the Shoah, and antisemitism (Alvén, 2024; Goldberg & Savenije, 2018; Kello, 2016; Savenije et al., 2019). Goldberg and Savenije (2018), for example, argued that teaching sensitive historical issues requires teachers to integrate sound disciplinary knowledge, expertise in facilitating classroom discussions, and sensitivity to the emotional and identity-related dimensions of learning.

With regard to antisemitism-critical education in and out of classrooms, discussions continue about the critical factors to successfully combat antisemitism (Bernstein, 2020; Chernivsky, 2025; Chernivsky & Lorenz-Sinai, 2026). Similar to teaching sensitive historical issues (Goldberg & Savenije, 2018), key components include, first of all, comprehensive and disciplinary knowledge of antisemitism and related areas, such as Jewish life, religion, culture, and history (Chernivsky, 2025; Kultusministerkonferenz, 2022; Zentralrat der Juden in Deutschland et al., 2025). Furthermore, appreciating Jewish perspectives and lived experiences, practicing empathy and establishing room to reflect on emotions are also considered to be crucial for effective antisemitism-critical education (e.g., Chernivsky, 2025; Frommer, 2025; Habig, 2025). Educators also need the competencies to recognize antisemitism in its diverse forms—ranging from overt hostility to subtle stereotypes—and to respond in a timely, informed, and pedagogically sound way (e.g., Ministerium für Kultus, Jugend und Sport Baden-Württemberg, 2019). Ultimately, antisemitism-critical education depends not only on knowledge and skills but also on the absence of antisemitic attitudes among educators (Bernstein, 2020).

However, whereas empirical research has advanced society's understanding of antisemitic attitudes, insights about knowledge, empathy, and competencies needed for effective antisemitism-critical education remain fragmented. *Antisemitic attitudes* are a latent disposition to respond negatively to people, locations, and institutions linked to Judaism and Jewish culture, or Israel, and include cognitive, emotional, and behavioral components (Zick et al., 2025). Since World War II antisemitic attitudes have been regularly assessed through large-scale surveys, focusing on individual prejudices among non-Jewish individuals (Anti-Defamation League, 2025; Chernivsky & Lorenz-Sinai, 2026; Zick et al., 2025). Research found that agreement with antisemitic statements is rising—or stable at a high level—and is not limited to political extremists but also appears in the center of society. Especially agreement with Israel-related antisemitism has increased in recent years. Zick et al. (2025) connected this trend to the Hamas attack on Israel on October 7, 2023, and Israel's response and current politics in the Arab-Israeli Conflict (Zick et al., 2025). For Jews in Israel, as well as in the diaspora, the attacks of 7 October 2023 perpetrated by Hamas and allied Palestinian actors constitute a profound historical rupture (Freedman & Hirsh, 2024). Since then, a growing body of research has examined the psychological impact of these terrorist attacks – and the radicalization of antisemitism ever since then—on Jewish populations both within Israel (Feingold et al., 2024; Groweiss et al., 2024; Levi-Belz et al., 2024) and across the Jewish diaspora (Bankier-Karp & Graham, 2024; Bar-Halpern & Wolfman, 2025; Mattson et al., 2024).

Regarding the assessment of *knowledge* about antisemitism, research remains considerably less developed than studies on antisemitic attitudes, despite evidence suggesting that knowledge plays a crucial role in enabling effective antisemitism-critical education (e.g., Bernstein, 2020). Given the complexity of discourses surrounding antisemitism (Chernivsky, 2025), a solid foundation of factual knowledge—particularly regarding key terms, historical concepts, and interrelated phenomena—is indispensable (Zentralrat der Juden in Deutschland et al., 2025). This importance was further highlighted by Niedick et al. (2026), who recently conducted a scoping review of 176 educational modules designed for antisemitism-critical education and found that 155 explicitly aimed to enhance participants' knowledge. This finding underscores the widespread recognition of knowledge as a foundational component of such initiatives. Knowledge refers to information that is stored relatively permanently in long-term memory (Renkl, 2014). The literature distinguishes two forms of knowledge: declarative and procedural knowledge: Declarative knowledge comprises factual knowledge and knowledge about concepts, single facts, and complex relationships ("knowing that"). Procedural knowledge, by contrast, concerns the know-how of processes and procedures ("knowing how"; De Jong & Ferguson-Hessler, 1996). Both forms are essential for antisemitism-critical education: e.g., factual, conceptual and relational knowledge about antisemitism, Jewish life, religion and culture, history of the Shoah, as well as procedural knowledge about where to turn for help and how to intervene. Professional knowledge is an essential part of professional competencies of teachers (Baumert & Kunter, 2006).

However, despite the recognition of knowledge to be essential for antisemitism-critical education, quantitative empirical research using standardized knowledge assessments remains rare. Only a limited number of studies employed validated instruments to measure knowledge (e.g., Alper et al., 2019; Claims Conference, 2025). For example, a large-scale survey conducted by the Claims Conference (2025) included a small set of questions assessing knowledge about the Shoah among approximately 1,000 young adults (aged 18–29) per country. These questions included identifying the number of Jews murdered during the Shoah and naming a concentration camp. The findings revealed significant gaps in basic historical knowledge among the participating youth. Similarly, Foster et al. (2015) conducted a representative survey of around 7,950 secondary school students in England, assessing their understanding of the Shoah. Whereas many students recognized key terms such as "Holocaust," the study found that most participants lacked a deeper understanding of historical context, core events, and essential facts about the Shoah (Foster et al., 2015).

In evaluating antisemitism prevention interventions, few studies have used standardized tests to assess knowledge, resulting in some uncertainty regarding the effectiveness of these interventions. Gordon et al. (2004) assessed knowledge about the Shoah using five multiple-choice items, including questions such as "What were the Nuremberg Laws?" and "Who was Oscar Schindler?" Geißler's (2008) dissertation is the only study to date to employ standardized instruments to measure knowledge across a broad range of topics. Test items spanned diverse content, from "How often is Shabbat observed?" to "What is the name of the national movement of the Jewish people since the late 19th century?" Notably, whereas the authors of many studies claimed to assess knowledge or report impacts of educational interventions on knowledge, these studies often relied on self-reported measures—participants rated their perceived knowledge before and after program participation (Müller, 2021; Niedick et al., 2026; Stefaniak et al., 2022; Stefaniak & Bilewicz, 2016). It is unclear how reliable the subjective reports on learning gains are; it is possible that they reflect shifts in confidence rather than actual knowledge acquisition.

Beyond attitudes and knowledge, several other factors are crucial for effective antisemitism-critical education, including self-reflection (e.g., Chernivsky, 2025) and empathy (e.g., Habig, 2025). Building on these foundational elements, we argue that *empathy* with individuals affected by antisemitism is a central, yet underdeveloped component of such education. In recent years, attention to empathy has grown significantly, particularly due to the inclusion of Jewish and non-Jewish individuals who have experienced antisemitism in large-scale research projects over the past 15 years—providing opportunities to share their lived experiences (e.g., Chernivsky & Lorenz-Sinai, 2026; Fischer et al., 2020). In the context of antisemitism prevention, empathy is not understood as mere compassion or emotional partisanship, but rather as a consciously reflective stance. The current research builds on an understanding of empathy drawing on Rogers (1959/2009): Empathy refers to the ability to comprehend another person's internal frame of reference as accurately as possible "as if" one were that person, while still maintaining one's own perspective and without identification or becoming emotionally entangled (Frommer, 2025). Scholars widely agree that emotional responsiveness, the capacity to understand, and the ability to appreciate the perspectives of those affected by antisemitism are essential components of effective antisemitism-critical education (Bernstein et al., 2022; Chernivsky & Lorenz-Sinai, 2021; Frommer, 2025; Grimm et al., 2025). Despite this consensus, Niedick et al. (2026) found that only 6 out of 176 educational modules explicitly aimed to foster empathy, and only 13 sought to promote a change in perspective. Taken together, these findings suggest that whereas empathy is recognized as a key component, its integration into educational interventions remains limited and underdeveloped.

Empirical research on empathy has produced many assessment instruments. For example, in a systematic review, Lima and Osório (2021) found 23 distinct empathy measurement tools. Despite the lack of a unified definition, researchers agree that empathy is multidimensional. Davis's (1983) Interpersonal Reactivity Index, for instance, measures four subcomponents: fantasy, empathic concern, perspective taking, and personal distress. In the understanding of Rogers (1959/2009) and Frommer (2025), only scoring high on personal distress, for example, would indicate failed empathy. In the context of antisemitism-critical education, Fischer et al. (2020) developed a survey to evaluate empathy with individuals affected by antisemitism. Respondents rated their understanding with statements indicating feelings of Jewish individuals. Related constructs appeared in other studies, such as religious-based perspective-taking (Schweitzer et al., 2017) and empathy in history education (e.g., Bartelds et al., 2023). Despite this growing body of research, the diversity of conceptualizations and assessment tools continues to complicate the development of a coherent understanding of empathy's role in effective

antisemitism-critical education. Furthermore, the lack of standardized, validated instruments tailored specifically to this context also hinders comparative analysis and the evaluation of educational interventions.

Teachers are required to recognize, name, and address antisemitism—*competencies* that are central to effective antisemitism-critical education (Zentralrat der Juden in Deutschland et al., 2025). Following Weinert (2001), competencies refer to cognitive abilities that individuals can acquire through learning, as well as the readiness and capacity to apply these skills in diverse and variable contexts. In the context of antisemitism-critical education, this means not only knowing about the "how to", but acting appropriately in a situation where antisemitism occurs and handling teaching situations sensitively. In their scoping review, Niedick et al. (2026) identified the objective "action strategies"—such as media literacy and fostering self-efficacy—on 27 occasions. These components are among the closest conceptual approximations to competencies found in the literature review by Niedick et al. (2026). Nevertheless, empirical research on competencies relevant to antisemitism-critical education remains extremely limited, particularly regarding the use of standardized measurement instruments. Niedick et al. identified only six empirical studies evaluating the effectiveness of educational interventions, none of which specifically assessed gains in competencies but in knowledge, attitudes or learning goals. Appropriate pedagogical action in the field of antisemitism-critical education is highly context-dependent and therefore difficult to put into standardized tasks that do not measure knowledge but competencies.

Taken together, these (three) domains (knowledge, empathy, competencies) are mutually reinforcing dimensions of teachers' professional responses to antisemitism. Research on teaching sensitive history has emphasized the interdependence of disciplinary knowledge, attention to learners' emotions and identities (empathy), and competencies (e.g., Goldberg & Savenije, 2018). Accordingly, the three domains are best understood as complementary and mutually dependent. Knowledge provides the foundation for recognizing and contextualizing antisemitism. Empathy supports understanding and context-sensitive engagement with those affected (Goldberg & Savenije, 2018). Both serve as prerequisites for competencies in the field of antisemitism-critical education (OSCE & ODIHR, 2020). At the same time, disciplinary knowledge provides the necessary framework for empathetic responses, ensuring that empathy is informed by historical understanding rather than detached from it (Frommer, 2025; Habig, 2025). Accordingly, effective responses to antisemitism depend on the integration of all three dimensions rather than on any single domain in isolation. Empirical research has shown how antisemitism develops and manifests in society. Yet, whereas teachers are crucial in preventing and addressing antisemitism, it is not clear what specific knowledge, empathy, and competencies are required for them to do so effectively. Despite the existence of various prevention programs and high expectations for teachers, the empirical understanding of these necessary qualities is limited. Therefore, to evaluate educational interventions or assess teachers' knowledge, empathy, and competencies, developing a valid and reliable quantitative measurement tool is an essential first step.

2.2 Standardized assessment instruments: What is needed?

A valid and reliable measurement instrument needs a solid theoretical foundation that clearly defines the constructs to be assessed (Haladyna et al., 2016; Priebe, 2024). Numerous educational resources—handouts, training modules, and background briefings (Niedick et al., 2026)—provide practical guidance and contextual information to help teachers combat antisemitism. However, a comprehensive, theoretically grounded framework specifying the knowledge, empathy, and competencies required for antisemitism-critical education has long been missing—and was the starting point of the EMPATHIA³ project (Empowering Police Officers and Teachers in Arguing Against Antisemitism). Its goal was to develop a core curriculum for trainers of future teachers and police officers, and to identify essential knowledge domains for preventing and addressing antisemitism (EMPATHIA³, 2026; Tikvah Institut, in press).

To ensure that a measurement instrument is suitable for large-scale surveys and effectiveness evaluations, several psychometric criteria must be met, for example objectivity, reliability, validity and economy (Haladyna et al., 2016). Objectivity means consistent and neutral test administration, scoring, and interpretation, regardless of who conducts or analyzes the assessment. This reduces personal bias and ensures that results are reproducible across different contexts and evaluators. Reliability is the reproducibility and consistency of measurement over time and across items. Reliability of a test instrument at one time point is commonly assessed using internal consistency measures, such as Cronbach's alpha. For example, measuring antisemitic attitudes often uses multiple items with different wording to result in a reliable measure-

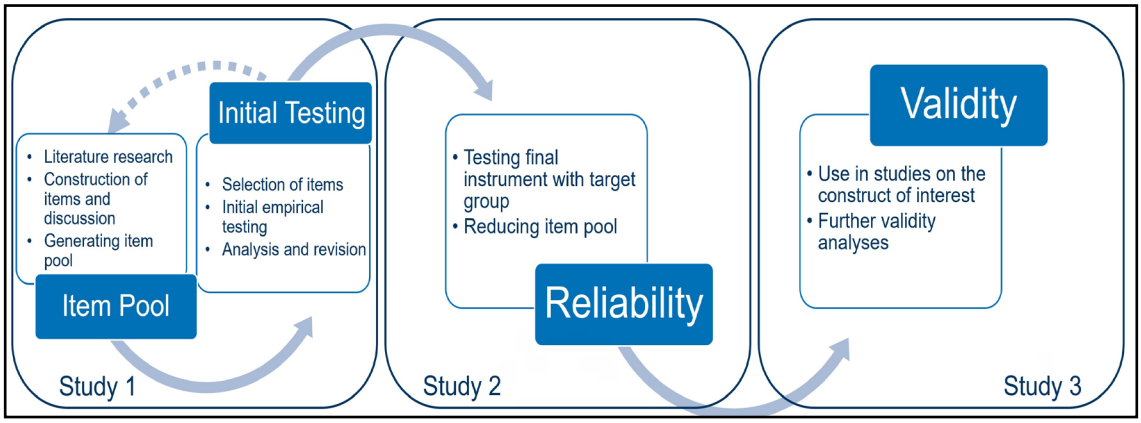
ment. Validity ensures that the instrument measures the intended construct and not something else. For instance, *content validity* is confirmed when subject-matter experts agree the items cover the construct of interest. *Construct validity* is assessed by examining the instrument's theoretical coherence with related constructs, such as the expected positive association between empathy and agreeableness. Economy refers to the resources needed to collect large amounts of data, test administration and interpretation (Haladyna et al., 2016).

3 Present studies

Empirical research in the domain of antisemitism-critical education for educators continues to lack a validated, comprehensive measurement instrument capable of assessing knowledge as the foundation to measure effective antisemitism-critical education. Furthermore, there is no instrument assessing empathy with those affected by antisemitism, and key competencies, which are relevant constructs for antisemitism-critical education (Zachrich & Grimm, in press). The present work contributes to the development of such measurement instruments in three steps. In Study 1, we developed a first set of items for the defined constructs through collaborative work with several partners, experts, and the target group, and conducted a small pilot study to gauge the feasibility and appropriateness of the items. In this pilot study, it was especially important to develop items that meet the standard of content validity. Study 2 focused on the reliability and validity of the measurement instruments. Here, the goal was to reduce the item pool to those items that meet psychometric quality criteria. Study 3 was conducted to demonstrate that the refined item pool would work in the target group and, hence, to further examine reliability and validity.

In the collaborative EMPATHIA³ project, experts from different fields worked together: Experts from empirical educational research and quantitative methodology were responsible for the test development process and worked closely together with team members of the EMPATHIA³ project in the fields of Judaism, History Education, Antisemitism, Israel, and Jewish life. The test development started with a comprehensive review of relevant literature and existing measurement instruments. The development of the core curriculum was a fundamental step in defining the constructs relevant to the test. After defining the relevant knowledge domains and additional constructs in focus, the project team began item development by drafting, discussing, and refining them. From the resulting item pool, a set of high-quality items was selected and tested in a pilot phase to empirically evaluate their performance. A pilot study is the initial empirical testing of the instrument (Haladyna et al., 2016). Results of the pilot study informed how to proceed, whether more items were needed or whether the items could already be used further (see Figure 1, Box "Study 1"). In the EMPATHIA³ project, a pilot study and think-aloud interviews with the target group were conducted during the initial test development phase of item construction and pilot studies. The item pool was then empirically tested in a validation study to provide further insights into the reliability and validity of the assessment instruments (see Figure 1, Box "Study 2"). Finally, the instruments were administered in a large study to assess the construct of interest, but in the present article, primarily to identify further implications for the instruments' validity (see Figure 1, Box "Study 3"). The test construction process thus followed the process described in the literature (e.g., Haladyna et al., 2016).

Figure 1: Process of test development in the EMPATHIA³-Project.



Note: Keywords in blue boxes indicate focus of the study.

4 Study 1

Study 1 describes the first phases of the test development process, including the construction of a first set of items. Item construction of knowledge and empathy was built on the core curriculum on antisemitism prevention. The *Foundations of Antisemitism Prevention – A Core Curriculum for Public Officials* (translated; Tikvah Institut, in press) provides a curriculum for antisemitism-critical education for prospective civil servants, including the following key content areas: "Antisemitism," "Jewish Life, Religion, and Culture," "Shoah," "Israel and the Arab-Israeli Conflict," and "Law". The curriculum concludes by discussing professional empathy as a pivotal element of antisemitism-critical education. This knowledge base served as the foundation for developing test items. Because effective antisemitism-critical education requires teachers to recognize antisemitism in its various forms and codes and address it, competencies in recognizing and addressing antisemitism adequately were additional constructs the test aimed to assess.

4.1 Method

4.1.1 Item construction

The core curriculum served as the main framework for item development. The goal was to develop items that assess not only crucial content knowledge across the different content areas but also empathy and professional competencies. To ensure content validity, team members who were responsible for writing the chapters of the core curriculum (according to the defined content areas) were designated as content experts for item development in their respective domains. For example, the authors of the chapter *Antisemitism* collaborated with the methods expert in a dedicated working group focusing on developing items related to this topic.

In the first phase, we conducted a comprehensive literature and resource review, drawing on existing measurement instruments (e.g., Alper et al., 2019; Fischer et al., 2020; Foster et al., 2015), qualitative studies (e.g., Bernstein, 2020), educational exercises from school text books and teaching handouts (e.g., Halbfas, 2009; Wetzel, 2006), and additional relevant materials. These included websites such as www.stopantisemitism.org, the short film Masel Tov Cocktail (Khaet & Paatsch, 2021), and educational online materials (Servicestelle für Antidiskriminierungsarbeit, Beratung bei Rassismus und Antisemitismus [SABRA], 2025)—that could be used as materials for item development. All materials informed systematic item construction.

All team members involved in item development participated in an initial item workshop, where principles of standardized test construction and item development were discussed, followed by a brainstorming session to generate items. Subsequently, item development proceeded in bilateral working groups, each consisting of the methods expert and a content expert. The resulting items were compiled and reviewed with the larger team, where all content experts provided feedback on items from other domains. At this stage, items were retained only if the team unanimously agreed on the topic's relevance and the correct answer was clearly identifiable. Items were excluded if they required excessive reading, were ambiguous, or addressed topics deemed too complex to be adequately assessed through standardized formats.

The draft test version was then divided by content area and sent to external experts for feedback. We first approached members of the advisory board of the EMPATHIA³ project; when they were unavailable, we contacted additional experts from the team's professional network, ensuring that everyone had demonstrated expertise in the relevant domain. In total, eight external experts provided feedback on various sections of the test. Concurrently, three think-aloud interviews were conducted with prospective teachers. Participants completed the test in an interview format and were asked to verbalize their thought processes while answering each item, allowing us to assess comprehension and potential sources of confusion.

Feedback from external experts and think-aloud interviews highlighted several critical concerns: the risk of reinforcing stereotypes through certain items, the need for clearer justification of correct answers (some of which relied on assumptions not explicitly stated), and the irrelevance of certain content to Jewish life in Germany. Based on this feedback, several items were substantially revised or removed entirely, particularly when consensus on a revised version could not be reached. Smaller revisions focused on wording, clarity of instructions, and definitions of key terms. After multiple rounds of iteration, expert feedback, and refinement, the first phase of test development yielded a pool of 91 knowledge items, 11 empathy items, and 29 competence items.

4.1.2 Construction principles, scoring and example items

The knowledge items were presented in several formats. The majority followed either a complex-multiple-choice format (one or more response options may be correct) or a standard multiple-choice format (exactly one answer is correct). Each item began with a brief stem (max. three lines) that posed the knowledge question, followed by an instruction indicating whether a single answer or one-or-more answers were required. The number of response options ranged from four to twelve. All stems and correct answers were derived directly from the core curriculum; distractors were either invented or based on common misconceptions about the topic. An example (multiple-choice item) was:

When did antijudaism develop? Only one answer is correct.

(1) With the raise and spread of Christianity
(2) With the sermons of the First Crusades
(3) With the emergence of racial ideology in the 19th century
(4) With the National Socialists' rise to power
(5) In a different time: _____

Note. The correct answer is (1).

A smaller set of items used alternative formats. For instance, one task presented twelve pictures of religious sites (e.g., the Western Wall, Al-Aqsa Mosque), objects (e.g., a communion cup), or symbols (e.g., the Star of David). Participants had to assign each picture to Judaism, Christianity, Islam, or "other religion." Another task on the history of Jerusalem listed several historical events that had to be placed into three chronological periods. A further scenario-based item asked participants to evaluate the legal status of different actions (e.g., "a person distributes stickers bearing swastikas" or "a teacher shows a portrait of Hitler during a lesson on National Socialism"). For each scenario respondents indicated whether the behavior constituted a criminal offense, not a criminal offense, or a criminal offense only in a specific context.

We decided not to include a "don't know" option, as research suggests this category may reflect personality traits such as uncertainty avoidance and is often selected even when participants have partial knowledge (e.g., Tourangeau et al., 2000). Items were scored as 1 only if all subitems (response options) were answered correctly, and 0 if at least one subitem was answered incorrectly.

Empathy-related items were administered as statements for participants to evaluate. Eleven self-developed items were inspired by Fischer et al. (2020), who conducted a survey on Jewish and non-Jewish perspectives on antisemitism. Item construction was guided by the understanding of empathy to comprehend the other's frame of reference without losing one's own (Frommer, 2025; Rogers, 1959/2009). Responses were recorded on a five-point scale (1 = *I don't understand that at all* to 5 = *I completely understand*). For example: "Many Jews are afraid of hostility and therefore wear religious symbols only in secret."

The competence items were also administered in either a complex multiple-choice or a multiple-choice format. The items differed in the way the stem was presented: some used short written scenarios, others employed caricatures or photographs. In the response options participants had to indicate whether the presented situation could be judged as antisemitic and, if so, why. A variety of materials was used to depict different contexts, for example: posts on the social media platform X, lyrics from German rap songs, or slogans displayed at a demonstration. These items were designed to capture the many, often subtle, manifestations of antisemitism that can be difficult to classify unequivocally. Items assessing adequate responses to antisemitic incidents were dropped before the pilot study because of ambiguity. Scoring followed the same 1 0 scheme applied to the knowledge items (1 = correct, 0 = incorrect).

4.1.3 Sample & design of pilot study

The pilot study took place in December 2024 and recruited participants through the mailing lists of the University of Tübingen. A total of N = 416 university students participated. The average age was M = 23.86 years (SD = 5.08); 69% identified as female, 28.2% as male, and 2.7% as diverse. 19% were teacher trainees. Among the remaining 81%, participants were enrolled in various subjects, from Humanities (29.2%) to Medicine and Health Sciences (12%).

To manage the large number of items in the study, the test was split into three versions, each targeting a completion time of about 30 minutes. After consenting, participants were randomly assigned to one of the versions (n_1 = 135, n_2 = 139, n_3 = 142). As an incentive, book vouchers (5 × 60€) were awarded by random draw.

4.2 Results and discussion

4.2.1 Knowledge

91 knowledge items were tested. Figure 2 depicts a sample item from the knowledge field of "Jewish Life, Religion, and Culture". The term "item" summarizes the responses to all subitems, indicating whether a participant scored correctly or incorrectly.

Figure 2: Sample item for the knowledge domain "Jewish Life, Religion, and Culture".

According to Jewish religious law (Halacha), a person is Jewish, if the person…

More than one answer may be correct.

(a) has a Jewish mother.

(b) was born in Israel.

(c) went through a conversion process.

(d) has a Jewish father.

(e) had the confirmation.

The correct answers of this item are (a) and (c).

Solution frequencies ranged from 0% to 100%. 65 items with solution frequencies between 10% and 90% showed acceptable item discrimination values. In contrast, items solved by more than 90% or fewer than 10% of participants lacked sufficient variance to differentiate between high- and low-knowledge individuals. Furthermore, 75 items exhibited discrimination values above .3. Regarding reliability, Cronbach's alpha ranged from α = .21 (M = 0.72, SD = 0.09, n = 23; Law) to α = .69 (M = 0.73, SD = 0.12, n = 34; Jewish Life, Religion, and Culture). For example, in the item shown in Figure 2, 96% responded correctly to (e) by not checking the subitem, but only 36% checked (c) correctly. Discrimination for this item ranged from 0.1 to 0.25, indicating the need for improvement. Notably, most of the items in the knowledge domain of "Antisemitism and Law" were excluded due to their difficulty and, therefore, lack of variance and low discrimination values.
To track changes in the knowledge tasks, adaptations were categorized as follows: 1 = *minor changes* (e.g., wording or task header, if meaning remains unchanged), 2 = *moderate changes* (e.g., one answer option added or removed), and 3 = *major changes* (e.g., multiple items altered in meaning). Based on the analysis of the pilot study and subsequent discussions with the expert group, 45 items were revised: 21 in category 1, 16 in category 2, and 8 in category 3.

4.2.2 Empathy

Eleven items were presented as statements for participants to evaluate. Internal consistency (as an indicator of reliability) of the Empathy scale was good, α = .89 (M = 3.21, SD = 0.51, n = 113). Example items are "Some Jewish people do not report antisemitic threats because they fear that law enforcement agencies will not take them seriously or that they will face even more problems" or "Many Jewish men and women perceive antisemitism as ever-present in their lives."

4.2.3 Competency

Twelve items assessing competencies in recognizing antisemitism were tested in the pilot study. Solution frequencies ranged from 9% to 100%. Of the 61 subitems, 25 had critical difficulty values below 10% or above 90%. 12 subitems showed discrimination values above the threshold of .30. Cronbach's alpha ranged from .46 (M = 0.76, SD = 0.12, n = 14; Test version 1) to .62 (M = 0.74, SD = 0.12, n = 26; Test version 2).

Throughout the item construction phase, many competence items were created, but most did not survive expert review. Specifically, some were discarded because they required excessive reading, measuring text comprehension rather than actual competencies in responding to antisemitism (e.g., Trautwein et al., 2017). Items using caricatures were also removed, as they risked reproducing antisemitic stereotypes—images are particularly memorable and can reinforce harmful associations. Furthermore, the item review process highlighted that antisemitic incidents in schools are rarely clear-cut, making it difficult to determine correct or incorrect responses.

Figure 3 shows a sample item for assessing competencies in recognizing antisemitism. Solution frequencies ranged from 62% (d; correct answer) to 89% (b; distractor), indicating that the subitems were of medium to easy difficulty. However, item discrimination was low ($0.07 < r < 0.39$), suggesting that these items cannot substantially contribute to assessing competencies in recognizing antisemitism in Test version 3. Despite this limitation, we decided to keep the item because it described a realistic scenario of antisemitism in school settings, whereas most of the other school-related competence tasks were dropped. Thus, the decision on this item exemplifies the complexity of balancing psychometric and content validity.

Figure 3: Sample item for the assessment of recognizing antisemitism in schools.

A student requests a leave of absence from school to observe the so-called "Al-Quds Day." The student explains to the teacher that this is a religious holiday and expresses the desire to participate in a demonstration, which he claims is an integral part of the observance. Should the teacher grant the student the day off?

Only one answer is correct.

(a) Yes. It is a traditional Muslim holiday for which the student must be granted a leave of absence.

(b) Yes. It is a traditional Jewish holiday for which the student must be granted a leave of absence.

(c) No. Students are not permitted to be excused from classes for religious holidays of non-Christian origin.

(d) No. It is not a regular religious holiday. The event was established by the Islamic Republic of Iran to assert an exclusive claim to Jerusalem for Muslims.

The correct answer of this item is (d).

Note. The item was developed by Häußermann (2024)

4.2.4 Summary study 1

The initial phases of test development—item construction, think-aloud protocols, and empirical testing in the pilot study—provided valuable insights for subsequent item refinement. Multiple expert discussions and feedback loops highlighted the challenges of designing knowledge, empathy and competence items in a highly complex research area. Empirical testing revealed that many knowledge items were too difficult for the target group, whereas a few were too easy. Moreover, item discrimination was low for many items, indicating the need for further improvements. The empathy scale showed satisfactory scale statistics. Notably, discussions implied that it was especially difficult to construct valid competence items. Nevertheless, because we worked together with several experts to construct and discuss the items, content validity was established for all items tested in the pilot study.

5 Study 2

Study 2 aimed to investigate the reliability and validity of an adapted and improved version of the measurement instruments. For the knowledge test in particular, the goal of Study 2 was to test a large number of items and to reduce the item pool to a set of items that reliably measure the constructs of interest. We tested the association between knowledge and field of study which may serve as a proxy for interest, and assumed that students of the study programs History, Religion, and Philosophy would score higher on the knowledge test than those of other academic programs. Regarding empathy, Study 2 focused on further approving the scale's reliability. The study was used to further work on and check the remaining competence items.

5.1 Method

5.1.1 Sample & design

Study 2 included prospective teachers and university students from various programs, who were recruited via email lists and the study authors' personal contacts. Data collection took place in July 2025. In total, N = 525 university students participated, with n = 293 enrolled in teacher education programs and n = 232 in other academic disciplines. Participants had a mean age of 24.4 years. Of the sample, 70.1% identified as female, 26.8% as male, and 3.1% as diverse. The study was conducted using personal devices.

5.1.2 Instruments

The instrument contained 94 items assessing knowledge, including 35 multiple-choice and 59 complex multiple-choice items. Empathy with those affected by antisemitism was assessed with the questionnaire tested in the pilot study. Competencies to recognize antisemitism were assessed using 11 items.

To efficiently assess a large pool of items, three test versions were used. After providing consent for data processing, participants received one of three versions of the instruments, each containing approximately 30 items across all knowledge domains. All participants completed background questions on their field of study, age, gender, and high school Grade Point Average (GPA). As an incentive, again, book vouchers (5 × 60€) were awarded by random draw.

5.2 Results & discussion

5.2.1 Knowledge

Solution frequencies and discrimination were examined at both item and subitem levels. Items with very low solution frequencies (< 10% of participants answered correctly) were reviewed in detail to identify extremely difficult distractors or correct options that were frequently omitted. This process helped identify items that even high-knowledge participants struggled to answer correctly. These items were excluded from further analysis to improve the instrument's reliability.

Item Response Theory (2PL) models were estimated to assess difficulty and discrimination parameters, and weighted likelihood estimate (WLE) person separation reliability (WLE PSR; Andrich, 1982) was used as a measure of reliability. Items with extreme difficulty parameters (a = ± 2), SE > 1, or low discrimination parameters (b < 0.5) were flagged for further review and potential exclusion. Following the removal of these items, the 2PL model was re-optimized, and WLEs (Warm, 1989) were computed for each participant. Finally, these scores were standardized and used in subsequent analyses of variable associations.

Table 1 presents the range of difficulty and discrimination parameters for the original set of items across the three test versions.

Table 1: Solution frequencies, discrimination of items and Cronbach's Alpha for the initial set of knowledge items in Study 2.

Test version	No of items	Solved (in %)		Discrimination of items		Cronbach's alpha
		Min	Max	Min	Max	
1	31	1	96	-0.08	0.46	0.68
2	36	1	97	0.06	0.50	0.74
3	27	1	91	0.06	0.41	0.71

Note. 123 < *N* < 147.

Analysis with 2PL models revealed the following (see Appendix A for detailed report of item discrimination and difficulties for each item): In Test version 1, 11 items were excluded due to high standard errors and low discrimination, leaving 17 items for the final 2PL model. Difficulty parameters ranged from −0.92 to 2.61 (0.15 < *SE* < 1.34), and discrimination parameters ranged from 0.4 to 1.61. The WLE PSR for the final model was 0.71. In Test version 2, 8 items were excluded because their difficulty estimates had large standard errors. The final 2PL model included 21 items, with difficulty parameters ranging from –1.98 to 2.52 (0.15 < *SE* < 0.89) and discrimination parameters from 0.21 to 0.68. Reliability was WLE PSR = 0.71. In Test version 3, 7 items were excluded, resulting in a final knowledge test of 18 items. Difficulty ranged from –1.75 to 2.51 (0.17 < *SE* < 0.80), and discrimination parameters ranged from 0.49 to 1.73. WLE PSR was 0.65. Reliability based on Item Response Theory is often lower than Cronbach's alpha because it takes into account that measurement precision varies for individuals with different latent trait levels. Overall, internal consistency was acceptable across all versions, discrimination values were adequate in the optimized models, and difficulty parameters covered a range of levels.

Table 2 shows the range of difficulty and discrimination parameters for the optimized set of items as well as the descriptive statistics for the knowledge sum score. On average, participants correctly answered fewer than 50% of the items.

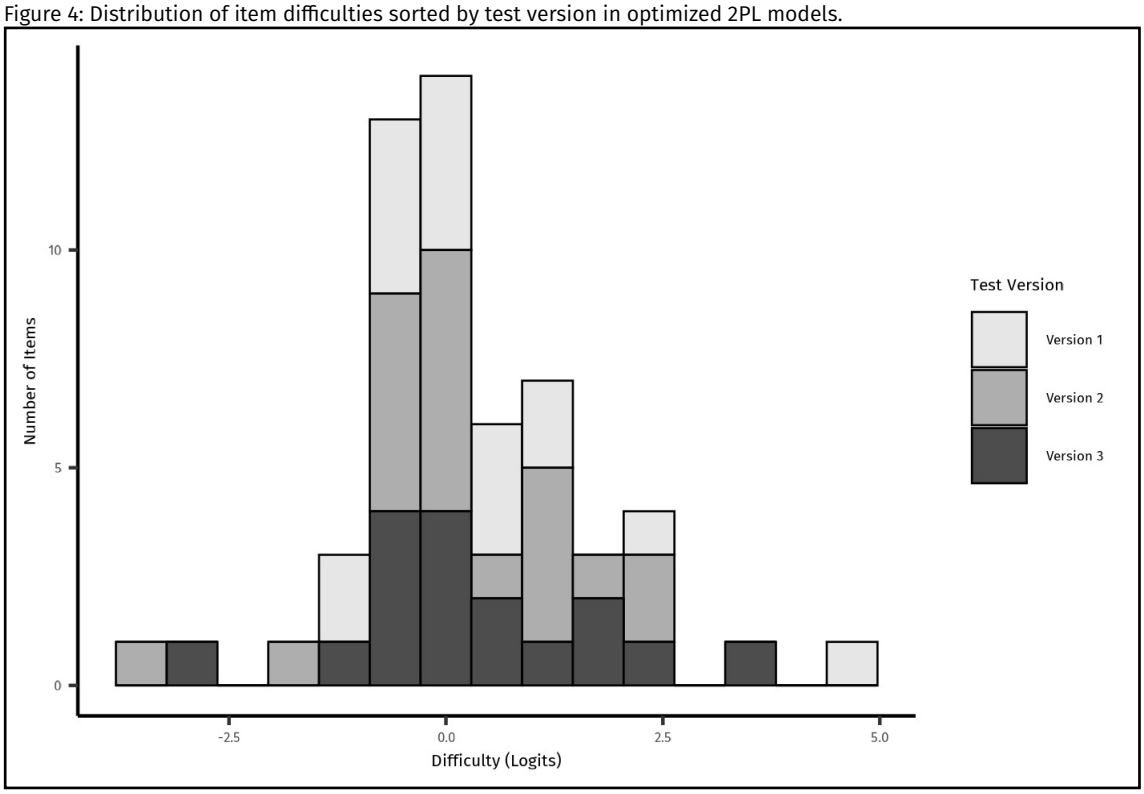
Table 2: Descriptives for knowledge scores with the final item set in study 2.

Test version	No of items	Solved (%)		r_{it}		Testscore				Cronbach's alpha
		Min	Max	Min	Max	*MScore*	*SDScore*	Min	Max	
1	17	6	73	0.14	0.50	7.41	3.53	0	16	0.74
2	21	13	92	0.19	0.52	9.34	4.26	0	21	0.76
3	18	7	87	0.20	0.40	7.48	3.61	0	16	0.70

Note. 138 < *N* < 148; r_{it} = discrimination of items.

Based on the three 2PL models, we developed an item pool with a wide range of difficulty levels, which served as the basis for item selection in subsequent studies. Figure 4 gives an overview on the distribution of difficulties across items and test versions. In this study, 15 items were revised: 8 in category 1 (minor changes, e.g., wording or task header), 5 in category 2 (moderate changes, e.g., addition or removal of one answer option), and 2 in category 3 (major changes, e.g., alterations in item meaning). The final item pool included 56 items, forming the foundation for Study 3.

Figure 4: Distribution of item difficulties sorted by test version in optimized 2PL models.



To further probe for the validity of the knowledge test items, we calculated correlations and analyses of variances (ANOVAs) using standardized WLE scores and personal background variables. Specifically, we hypothesized that students in Humanities (e.g., History, Philosophy, Religion) would score higher than those in other programs because the study field may represent a field of interest, and interest is a strong predictor for knowledge acquisition (e.g., Krapp, 1999). This hypothesis was supported by statistically significant mean differences (M_{Diff}= 0.36, t = –2.856, df = 193.9, p = .005). Moreover, high school GPAs often serve as indicators of general skills and as predictors for later academic success, i.e., knowledge acquisition (e.g., Steedle & Way, 2024). Although better grades tended to be related to higher knowledge, the association was not statistically significant (r = –.10, p = .060).

5.2.2 Empathy

Discrimination values for the empathy items ranged from .51 to .74. We checked the potential for a shortened version to improve test economy. The shortened scale, with four items removed, showed good internal consistency (α = 0.87). The correlation between the shortened and original scales was r = .97, indicating that the shorter version is a suitable substitute. Seven items were therefore retained for further research. The scale mean was 3.93 (SD = 0.82), N = 525.

Empathy statistically significantly correlated with the standardized WLE knowledge score (r = .12, p = .014). There were no statistically significant correlations between empathy and age, number of semesters at university, or high school GPA. We also did not find a statistically significant correlation between empathy and students' study field. We found a statistically significant association between gender and empathy, indicating that people identified as female scored higher on empathy than those identified as male (M_{Diff} = 0.29, t = 3.05, df = 161.41, p = .003).

5.2.3 Competency

Due to the small number of competence items—three or four per test version—analyses were conducted descriptively, examining means, standard deviations, and correlations with WLE knowledge scores. Means ranged from 0.24 to 0.71 (0.43 < SD < 0.50); detailed descriptive statistics for each item are provided in Appendix B. We found substantial positive correlations between the standardized WLE knowledge score and two items: Item "From Berlin to Gaza–Yallah Intifada" (Appendix B, C3), r = .36, p < .001, and Item "Al-Quds Day" (see Figure 3; Appendix B, C4), r = .35, p < .001. For all other competence items, correlations were not substantial or statistically

significant, ranging from $r = -.11$ to .13 ($ps > .114$). These results support the notion that assessing competencies in antisemitism-critical education remains challenging. Based on these findings, we decided to retain four competence items for further research.

5.2.4 Summary study 2

In Study 2, many knowledge items were tested and analyzed using Item Response Theory. Item difficulty and discrimination informed decisions about whether to adapt or exclude items from further analysis. Reliability scores for the knowledge tests across the three test versions were acceptable. Mean differences in knowledge test scores between students studying Humanities compared to those studying other academic programs, which served as a proxy for interest in the topics of the knowledge test, provided initial support the test's construct validity. Empathy with those affected by antisemitism showed good reliability. Regarding measuring competencies, little progress was made because the project team only agreed on a few items for testing. Items of the knowledge test were categorized as easy, medium, or difficult and sorted by knowledge domain, forming a foundation for constructing tests in future research. This categorization guided item selection for Study 3. Additionally, all items underwent a final review in an expert discussion to further strengthen the content validity of the knowledge items.

6 Study 3

The primary aim of Study 3 was to examine how prospective teachers perform on the newly developed instruments and to gain further insights into the instrument's reliability and validity. We specifically investigated associations between knowledge and personal background variables, self-reported knowledge, interest, and self-reported competencies. Mean differences across participants in different study fields again served as indicators of construct validity, because study programs may serve as a proxy for interest. Thus, we assumed that students in study programs more closely related to the topic of antisemitism and neighboring knowledge domains would score higher than students in other programs. Additionally, we examined the association between empathy with those affected by antisemitism and a validated empathy inventory, as well as personal background variables. The competence items we retained from Studies 1 and 2 were used to assess their performance in a larger sample of the target group.

6.1 Method

6.1.1 Sample & design

Two strategies were used to recruit prospective teachers in North Rhine-Westphalia (NRW) to participate in Study 3. First, the study team contacted the managing directors of all teacher education departments at universities via email and phone, requesting that they share the call for participation with their mailing lists. Whereas some universities lacked comprehensive lists for all teacher education students, others had policies restricting the use of mailing lists for research recruitment. Nevertheless, all addressed 11 universities in NRW supported recruitment by either forwarding the call or posting it on at least one study platform, such as Moodle. The second strategy involved contacting lecturers of large lectures on teaching and learning which are part of Teacher Training programs across all universities, asking them to include a call for participation in their lectures. Contact was made with 109 lecturers.

Participants received a compensation of €5 upon full completion of the study. Data collection took place from October 24 to November 30, 2025. A total of $N = 3{,}721$ university students participated. The mean age was 23.27 years ($SD = 4.12$). 51.6% were in semesters 1 to 6, 27.7% in semesters 7 to 10, 17% in semesters 11 to 15, and 3.7% in semester 16 or higher. In total, 814 participants were enrolled in Humanities (e.g., History, Religion, Philosophy), whereas 1,790 studied other programs. The mean high school GPA was 2.19 ($SD = 2.2$). Of all participants, 69.9% identified as female, 28.4% as male, and 1.7% as diverse.

6.1.2 Instruments

Knowledge, empathy, competencies. The results of the optimized 2PL model from Study 2 guided the selection of items for Study 3. Based on these results, experts aimed to ensure each knowledge domain was represented by items of varied difficulty. Experts selected items using a rule of thumb to include approximately one easy, one difficult, and two medium-difficulty items per knowledge domain. In total, 30 knowledge items were administered—13 in multiple-choice format and 17 in complex-multiple-choice format. Four competence items were included. Empathy with those affected by antisemitism was assessed using the shortened scale (7 items).

Self-reports. Additionally, we assessed self-reported knowledge, and self-reported interest. For each knowledge domain participants rated their knowledge and interest on a five-point scale: 1 = *no knowledge/not at all interested,* to 5 = *a lot of knowledge/very much interested*. Participants also rated their perceived ability to recognize and address antisemitism (1 = *no skills* at all to 5 = *quite some skills*).

Problem awareness. We asked how participants feel about addressing antisemitism in schools and how they evaluate the need for prevention programs, using two questions adapted from Bernstein (2020). The first question asked how participants feel when it comes to the topic of dealing with antisemitism in school. Response categories were nominal and included (1) "overwhelmed," (2) "not trained," (3) "indifferent" (with two distinct reasons: low perceived prevalence of antisemitism in schools, or not intending to teach a subject typically addressing the topic, such as history, politics, religion, or German), (4) "annoyed" (due to perceived overemphasis of antisemitism), (5) "well-prepared," and an open (6) "other" option. We summarized responses indicating an awareness of the problem (e.g., 'overwhelmed') and responses indicating indifference or annoyance (e.g., 'annoyed'). We assumed that participants who indicated an awareness of the problem would score higher on knowledge and empathy than those who did not. The second question tapped participants' beliefs about whether preventive programs to combat antisemitism are necessary for teachers in schools. Response categories were nominal; participants selected one of seven statements, ranging in content from "such programs are absolutely necessary, as racism and antisemitism also occur among teachers" to "there is no need, as there are hardly any Jewish students in German schools," with the seventh being an open-ended "other" option. Similarly, we summarized responses indicating that there is no need for preventive programs and those indicating that such programs are necessary. We assumed that participants who saw a need for preventive programs would score higher on knowledge and empathy than those who did not.

General empathy. General empathy was assessed using the German version of the Interpersonal Reactivity Index (SPF; Davis, 1983; Paulus, 2009). This instrument distinguishes four subscales—empathic concern, perspective-taking, personal distress, and fantasy—each measured by four items. Participants indicated how often each statement applied to them on a five-point scale from 1 = *never* to 5 = *always*. A sample item from the empathic concern subscale is: "I feel a deep sense of compassion for people who are less fortunate than I am."

6.2 Results and discussion

For Study 3, we performed analyses similar to those in Study 2, including descriptive statistics at the item level, focusing on solution frequencies and discrimination values. We examined response patterns to identify overly difficult subitems and estimated a 2PL model at the item level. As in Study 2, model thresholds helped identify ineffective items. Using psychometric criteria, we developed an optimized version of the test and evaluated WLE scores from this version in relation to self-reported knowledge, interest, and competencies.

6.2.1 Dropout rate & missing responses

The dropout rate in Study 3 was 30.23%. Hence, 1,125 participants dropped out before completing the test, whereas N = 2,596 completed the test to the final item. The item nonresponse rate for knowledge items ranged from 20% to 31%. For empathy, item non-response ranged from 13% to 15%. For competence items, it ranged from 30% to 31%. Missing data were handled using pairwise deletion; sample sizes therefore varied slightly across analyses.

6.2.2 Descriptive statistics

Appendix C provides a detailed table of item statistics. Solution frequencies for the knowledge items ranged from 1% ("criminal acts", Appendix C, K21) to 72% ("Auschwitz-Birkenau", Appendix C, K27). Cronbach's alpha for the knowledge test was α = 0.78. Four items showed discrimination values below 0.2, indicating weak discrimination. The mean knowledge sum score was *M* = 8.37 (*SD* = 4.93), indicating that the test was relatively difficult with a maximum score of 30.

Table 3: Descriptive statistics and correlations for general empathy and empathy with those affected by antisemitism in Study 3.

	M	*SD*	(2)	(3)	(4)	(5)
(1) Empathetic concern	3.85	0.59	0.44	0.25	0.37	0.24
(2) Fantasy	3.43	0.72		0.25	0.27	0.18
(3) Personal distress	2.85	0.73			-0.02	0.12
(4) Perspective taking	3.66	0.61				0.15
(5) Empathy with those affected	3.66	0.85				

Note. 3,155 < *N* < 3,385.

Internal consistency of empathy with those affected by antisemitism was α = 0.88, *M* = 3.66, *SD* = 0.85. Descriptive statistics of general empathy's subfacets and empathy with those affected are shown in Table 3.

Table 4: Descriptives for self-reported knowledge, interest, and competencies in Study 3.

Construct	Domain	*M*	*SD*
Self-reported knowledge	Judaism and Jewish life	2.76	0.90
	Antisemitism	3.34	0.92
	History of National Socialism & Holocaust	3.95	0.82
	Israel & Arab-Israeli conflict	2.73	1.11
Self-reported interest	Judaism and Jewish life	2.82	1.03
	Antisemitism	3.47	1.01
	History of National Socialism & Holocaust	3.98	0.93
	Israel & Arab-Israeli conflict	3.25	1.15
Self-reported competencies	Recognizing antisemitism	3.33	0.94
	Addressing antisemitism	3.00	0.99

Note. 3,311 < *N* < 3,327. Item refers to exact wording of assessment (translated). Discrepancies between wording and titles of chapters in the core curriculum are due to the progress of the project. All items were answered from 1 = *no knowledge at all/no interest at all/weak skills* to 5 = *a lot of knowledge/a lot of interest/strong skills*.

Descriptive statistics and correlations for the four competence tasks are depicted in Appendix D. Cronbach's alpha was α = .30; item intercorrelations ranged from 0.02 to 0.20. Accordingly, no scale score was computed. Self-reported interest, knowledge, and competencies of recognizing and addressing antisemitism are depicted in Table 4.

6.2.3 Test statistics for knowledge items

The 2PL model revealed one item ("criminal acts", Appendix C, K21) with a critical discrimination value, which was therefore excluded from further analysis. The final 2PL model included 29 items, with difficulty parameters ranging from –1.11 to 3.51 (0.05 < *SE* < 0.34), and discrimination parameters ranging from 0.52 to 1.72 (0.05 < *SE* < 0.10). WLE PSR was 0.66. Model fit was good (*SRMR* = 0.024). In total, the final knowledge test included 3 easy, 5 medium, and 20 difficult items. Estimates of difficulty, discrimination, and standard errors for each item are provided in Appendix C.

6.2.4 Validation analyses

In the next step, we analyzed the associations between knowledge and empathy and relevant background variables. We assumed that a better high school GPA would be associated with higher knowledge scores. For age and the number of semesters at university, we assumed positive associations with knowledge. We did not assume associations of high school GPA, age, and semesters with empathy. We expected students in the Humanities to score higher on the knowledge test, because their fields of study may reflect greater interest in the topic. We did not expect any differences between study fields in empathy. We expected people who identified as female to score higher on empathy.

We found a small negative correlation between WLE scores and high school GPA (r = –.14, p < .001), and medium positive correlations with age (r = .24, p < .001) and number of semesters of study (r = .25, p < .001). The correlations between these variables and empathy were low, although the correlations with high school GPA and semesters were statistically significant. Results of the mean-difference analyses supported the assumption that students in the Humanities scored higher on knowledge, and no differences were found in empathy. However, unexpectedly, we found statistically significant gender differences in mean scores for knowledge and empathy, indicating that people who identified as male scored higher on both instruments. Empathy and knowledge were moderately correlated (r = .15, p < .001).

We expected that participants who indicated awareness of the problem of antisemitism in schools would score higher on knowledge and empathy. 90% of the participants indicated an awareness of the problem. Similarly, 86% indicated that they see the need for preventive programs to combat antisemitism. T-tests revealed statistically significant group differences for the investigated variables: Differences for knowledge by problem awareness (M_{Diff} = -0.82, t = -11.93, df = 303; p < .001) and need for preventive programs (M_{Diff} = -0.46, t = -7.91, df = 517.32; p <.001). Hence, people with higher knowledge statistically significantly showed more awareness and need for preventive programs than those with lower knowledge scores. A similar result pattern was found for differences in empathy with those affected by antisemitism: Differences for empathy by problem awareness (M_{Diff} = -0.68, t = -10.92, df = 278.68, p < .001); and need for preventive programs (M_{Diff} = -0.75, t = -15.34, df =454.92, p < .001). As hypothesized, the awareness of the problem of antisemitism in schools and the need for preventive programs is statistically significantly related to empathy with those affected by antisemitism. These results provide hints about the validity of the knowledge test and the empathy scale.

Correlations between WLE test scores and self-reported knowledge ranged from r = .12 (Israel and the Arab-Israeli Conflict) to r = .24 (History of National Socialism and the Shoah); all *ps* < .001. These small-to-moderate correlations indicate a limited association between overall test performance and self-reported knowledge, suggesting a discrepancy between self-evaluation and standardized assessment and highlighting the need for objective measurement tools. Correlations between test scores and self-reported interest were slightly higher, ranging from r = .16 (Israel and the Arab-Israeli Conflict) to r = .26 (Judaism and Jewish Life); all *ps* < .001. Knowledge scores were moderately correlated with self-reported competencies to recognize antisemitism (r = .26, p < .001). However, there was no significant correlation between WLE test scores and self-reported competencies to address antisemitism (r = .05, p = .008).

To assess empathy, we investigated correlations between the SPF subscales and empathy with those affected (see Table 3). Correlations were small to medium in size. These results support the assumption that the scale developed in the EMPATHIA³ project may measure a distinct aspect of empathy, rather than general empathy, but is related to empathetic feelings.

6.2.5 Summary study 3

In Study 3, 30 knowledge items were administered to assess knowledge. Item Response Theory analyses led to the exclusion of one item, resulting in a reliable assessment instrument. Furthermore, correlational analyses provide further support for validity, indicating that a first version of a standardized instrument is now available to measure relevant knowledge in antisemitism-critical education. In addition, empathy showed very good reliability, and correlations with the general empathy's subscales indicated that the association with empathetic concern is of medium size. Furthermore, empathy with those affected by antisemitism seems to measure a distinct aspect of empathy which general empathy inventories do not assess. Regarding competencies, only four items were administered, which exhibited low inter-correlations and internal consistency; thus, further effort is needed to assess competencies relevant to antisemitism-critical education.

7 General discussion

The studies yielded an item pool with a range of item difficulties to assess knowledge that is relevant to antisemitism-critical education and that can be used in future studies. Analyses revealed good internal consistency as well as content, and construct validity. In addition to the knowledge measures, the studies allowed to develop a scale with sound psychometric qualities which assesses empathy with those affected by antisemitism as a specific part of empathy, which is not yet included in general empathy assessment tools. This seven-item inventory is ready to use in quantitative-empirical research projects. Conversely, our items assessing competencies in recognizing and addressing antisemitism did not meet psychometric quality criteria; hence, assessment of competencies remains challenging and requires further research and refinement. Several efforts exist to support antisemitism-critical education and to prepare teachers to address antisemitism effectively (e.g., Niedick et al., 2026). However, until now, almost no studies on the effectiveness of such programs exist using standardized assessment tools and therefore are limited in their explanatory power (Müller, 2021; Zachrich & Grimm, in press). With the standardized knowledge test items and the empathy scale presented in this article, sound methodological research is now possible to investigate the effects of such programs on teachers' knowledge and empathy with those affected by antisemitism. Additionally, the instrument can be used to conduct assessment studies of teachers' knowledge and to identify specific needs for further education. In line with teaching sensitive historical issues, standardized tasks provide the opportunity for future research to examine the associations among foundational knowledge, sensitivity to emotions, and crucial competencies.

Results of the knowledge test raise the question of how knowledge standards should be defined and which items are essential for meeting them. Many items were difficult to solve, indicating that a lot more effort in the teacher training programs is needed. Regarding empathy, it is only one of several factors considered crucial for effective antisemitism-critical education. Understanding antisemitic motivations, openness to self-reflection, tolerance for ambiguity, and self-efficacy are also important. To advance research and practice, researchers and practitioners should collaborate to establish a comprehensive framework describing all relevant factors in antisemitism-critical education. This shared foundation will support the development and refinement of future measurement instruments. Furthermore, there should be clear distinctions between knowledge and competencies, as it is sometimes difficult to differentiate between competencies and procedural knowledge. The modest performance of the test in measuring the competencies that teachers need to foster antisemitism critical education underscores both the complexity of these competencies and the limitations of standardized test formats, as also found in previous research in other domains (e.g., Trautwein et al., 2017). Whereas factual knowledge can be captured adequately with short stem items, the ability to recognize and respond to antisemitism demands richer contextual information. Slogans, images, and brief scenarios, for example, require additional background to allow a reliable judgement of their antisemitic

content. Moreover, without a comprehensive conceptual framework and clear procedural guidelines, the evaluation of antisemitic incidents inevitably involves a normative component, making it difficult to assess such competencies with a purely objective, multiple choice format. Future research should therefore broaden its approach and consider alternative measurement tools for these competencies.

In summary, this article provides insights into the progress and challenges of assessing factors crucial to effective antisemitism-critical education. It presents a first version of test instruments for knowledge and empathy that meet psychometric quality standards and can be used to evaluate intervention studies. The test can be used to gain a deeper understanding of knowledge and empathy of prospective teachers in general and history teachers specifically, to identify needs for further education. The article also highlights the need for ongoing discussion of knowledge, standards, and competence levels, as well as the development of a comprehensive framework for antisemitism-critical education. Building on these needs, additional research is necessary to further develop measurement instruments and conduct studies that investigate the effectiveness of educational interventions fostering antisemitism-critical education.

To cite this article

Zachrich, L., Baier, J., Beck, V., Berger, D., Brauch, N., Frommer, J.-A., Grimm, M., Habig, J., Hauch, D. Z., Jahn, S. J., Mäder, J.-M., Mierwald, M., Netzer, K., Vosding, S., & Trautwein, U. (2026) Towards psychometrically sound assessments of knowledge, empathy and competencies in antisemitism-critical education: Progress and challenges. *Historical Thinking, Culture, and Education,* 4(1). https://doi.org/10.12685/htce.2222

Peer review

This article has been peer reviewed through the journal's standard double-blind peer review, where both the reviewers and authors are anonymized during review.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Grants

This work was supported by the Federal Ministry of Research, Technology and Space, Germany, under Grant 01UG2150D.

Ethical statement

The ethics committee of the Faculty of Social Sciences of the University of Tübingen granted ethical approval for the studies (ID: A2.5.4-379_hb; A2.5.4.-415_hb). Generative AI was used for language editing and proofreading to ensure linguistic correctness. The final manuscript was thoroughly reviewed and approved by the authors.

ORCID iD

Lisa Zachrich https://orcid.org/0000-0003-3037-986X

Nicola Brauch https://orcid.org/0000-0002-1463-4085

Johannah-Margarethe Mäder https://orcid.org/0000-0003-3090-3107

Katharina Netzer https://orcid.org/0009-0005-3139-5257

Marc Grimm https://orcid.org/0000-0002-5706-1053

Marcel Mierwald https://orcid.org/0000-0003-3530-4728

Sarah Jadwiga Jahn https://orcid.org/0009-0008-6340-4957

Ulrich Trautwein https://orcid.org/0000-0003-0647-0057

Volker Beck https://orcid.org/0000-0002-5664-1806

References

Alper, B. A., Smith, G. A., Schiller, A., & Nolan, H. (2019). *What Americans know about religion.* Pew Research Center.

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR 20 index, and the Guttman scale response pattern. *Education Research and Perspectives, 9*(1), 95–104. https://www.rasch.org/erp7.htm

Anti-Defamation League. (2025). *The ADL Global 100: Index of antisemitism.* https://www.adl.org/adl-global-100-index-antisemitism

Alvén, F. (2024). Controversial issues in history teaching. *Journal of Curriculum Studies, 56*(5), 537–553. https://doi.org/10.1080/00220272.2024.2322502

Bankier-Karp, A., & Graham, D. (2024). Surrounded by darkness, enfolded in light: Factors influencing the mental health of Australian Jews in the October 7 aftermath. *Contemporary Jewry,* 44(4), 903–935.

Bar-Halpern, M., & Wolfman, J. (2025). Traumatic invalidation in the Jewish community after October 7. *Journal of Human Behavior in the Social Environment, 35*(8), 1366–1393. https://doi.org/10.1080/10911359.2025.2503441

Bartelds, H., Savenije, G. M., Van Drie, J., & Van Boxtel, C. (2023). Using eyewitnesses to promote students' understanding of empathy in the history classroom. *The Journal of Social Studies Research, 47*(2), 129–144. https://doi.org/10.1016/j.jssr.2022.12.001

Bauer, U. (2021). Mit Bildung gegen das kulturelle Gedächtnis eines globalen Judenhasses—Geht das? In M. Grimm & S. Müller (Eds.), *Bildung gegen Antisemitismus* (pp. 21–43). Wochenschau.

Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520. https://doi.org/10.1007/s11618-006-0165-2

Bernstein, J. (2020). *Antisemitismus an Schulen in Deutschland—Analysen, Befunde, Handlungsoptionen.* Bundeszentrale für politische Bildung.

Bernstein, J., Grimm, M., & Müller, S. (2022). Addressing antisemitism in Germany: Challenges and possibilities in society, school, and education. *Journal of Contemporary Antisemitism, 5*(2), 29–42. https://doi.org/10.26613/jca.5.2.114

Chernivsky, M. (2025). Qualitätsmerkmale antisemitismuskritischer Bildung. *Journal für politische Bildung, 15*(1), 50–54.

Chernivsky, M., & Lorenz-Sinai, F. (2021). *Antisemitismus im Kontext Schule: Deutungen und Umgangsweisen von Lehrer*innen an Berliner Schulen.* Kompetenzzentrum für Prävention und Empowerment.

Chernivsky, M., & Lorenz-Sinai, F. (2026). *Bundesweite Studie zu den Auswirkungen des terroristischen Anschlags am 7. Oktober 2023 auf jüdische und israelische Communities in Deutschland—Zwischenbericht.* Kompetenzzentrum antisemitismuskritische Bildung & Forschung. https://archive.jpr.org.uk/download?id=24719

Claims Conference. (2025). *The first-ever-8-country Holocaust knowledge and awareness index.* https://www.claimscon.org/country-survey/

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

De Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105–113. https://doi.org/10.1207/s15326985ep3102_2

EMPATHIA³. (2026). *Empathia³.* https://empathia3.de/

Feingold, D., Neria, Y., & Bitan, D. T. (2024). PTSD, distress and substance use in the aftermath of October 7th, 2023, terror attacks in Southern Israel. *Journal of Psychiatric Research, 174*, 153–158. https://doi.org/10.1016/j.jpsychires.2024.04.022

Fischer, R., Harig, J., Holler, M., & Zwilling, C. (2020). *Mehrfachnennungen möglich.* Bildung in Widerspruch e.V.

Foster, S., Pettigrew, A., Pearce, A., Hale, R., Burgess, A., Salmons, P., & Lenga, R.-A. (2015). *What do students know and understand about the Holocaust? Evidence from English secondary schools.* Centre for Holocaust Education, UCL.

Freedman, R., & Hirsh, D. (Eds.). (2024). *Responses to 7 October: Antisemitic discourse.* Routledge.

Frommer, J.-A. (2025). *Professionalisierung durch Empathie. Ein selbstreflexiver Ansatz antisemitismuskritischer Bildung in öffentlichen Institutionen.* Universitätsbibliothek der Ruhr-Universität Bochum. https://doi.org/10.46586/E3.456

Geißler, N. (2008). *Konzeption und Evaluation eines Unterrichtsprogramms zur Prävention und Verminderung von Antisemitismus* [Doctoral dissertation, Justus-Liebig-Universität Gießen].

Goldberg, T., & Savenije, G. M. (2018). Teaching controversial historical issues. In S. A. Metzger & L. McArthur Harris (Eds.), *The Wiley international handbook of history teaching and learning* (pp. 503–526). Wiley-Blackwell.

Gordon, S. B., Simon, C. A., & Weinberg, L. (2004). The effects of Holocaust education on students' level of antisemitism. *Educational Research Quarterly, 27*(3), 58–71.

Grimm, M., Beer, F., & Kranz, D. (2025). *A safe school for Jewish pupils: A guide to antisemitism-critical school development in 35 questions and answers.* Universitätsbibliothek der Ruhr-Universität Bochum. https://doi.org/10.46586/E3.417

Groweiss, Y., Blank, C., Hamdan, S., Neria, Y., & Levi-Belz, Y. (2024). The mental health impact of the October 7th terror attack on Jews and Arabs in Israel: A nationwide prospective study. *Psychiatry Research,* 337. https://doi.org/10.1016/j.psychres.2024.115973

Habig, J. (2025). *Historisch denken – Emotionen „verstehen". Geschichtsbewusste Empathie als eine Komponente historischer Orientierung im Kontext des Antisemitismus der Gegenwart.* Universitätsbibliothek der Ruhr-Universität Bochum. https://doi.org/10.46586/E3.394

Haladyna, T. M., Lane, S., & Raymond, M. R. (Eds.). (2016). *Handbook of test development* (2nd ed.). Routledge. https://doi.org/10.4324/9780203102961

Halbfas, H. (Ed.). (2009). *Religionsbuch für das 5./6. Schuljahr.* Patmos.

Häußermann, N. (2024). *Islamischer Antisemitismus: Items zum Forschungsprojekt „Empathia3"* [Master's thesis, Universität Tübingen].

International Holocaust Remembrance Alliance. (2016, May 26). *IHRA non-legally binding working definition of antisemitism.* https://holocaustremembrance.com/wp-content/uploads/2024/01/IHRA-non-legally-binding-working-definition-of-antisemitism-1.pdf

Kello, K. (2016). Sensitive and controversial issues in the classroom: teaching history in a divided society. *Teachers and Teaching – theory and practice, 22*(1), 35–53. http://dx.doi.org/10.1080/13540602.2015.1023027

Khaet, A., & Paatsch, M. (Directors). (2021). *Masel Tov Cocktail* [Short film]. Filmakademie Baden-Württemberg.

Krapp, A. (1999). Interest, motivation and learning: An educational-psychological perspective. *European Journal of Psychology of Education, 14*(1), 23–40.

Kultusministerkonferenz. (2022). *Standards für die Lehrerbildung: Bildungswissenschaften* [Beschluss]. https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf

Levi-Belz, Y., Groweiss, Y., Blank, C., & Neria, Y. (2024). PTSD, depression, and anxiety after the October 7, 2023 attack in Israel: A nationwide prospective study. *eClinicalMedicine, 68.* https://doi.org/10.1016/j.eclinm.2023.102418

Lima, F. F., & Osório, F. L. (2021). Empathy: Assessment instruments and psychometric quality – a systematic literature review with a meta-analysis of the past ten years. *Frontiers in Psychology, 12.* https://doi.org/10.3389/fpsyg.2021.781346

Mattson, C., Andersson Mamlros, R., & Sager, M. (Eds.). (2024). *Antisemitism in Sweden after 7 October* (Report No. 13). The Segerstedt Institute. https://www.gu.se/en/segerstedt-Institute/publications/si-report-series

Ministerium für Kultus, Jugend und Sport Baden-Württemberg. (2019). *Wahrnehmen—Benennen—Handeln: Handreichung zum Umgang mit Antisemitismus an Schulen.*

Müller, S. (2021). Antisemitismusprävention als Bildungserfahrung: Wenn Wissen und Reflexion vor Ressentiments schützen sollen. In M. Grimm & S. Müller (Eds.), *Bildung gegen Antisemitismus* (pp. 214–231). Wochenschau.

Niedick, J., Juang, L. P., & Bitter, L. (2026). Education against antisemitism. A scoping review. *Zeitschrift für Bildungsforschung,* 1–7. https://doi.org/10.1007/s35834-025-00529-x

OSCE & ODIHR. (2020). *Addressing anti-Semitism in schools: Training curriculum for secondary education teachers.* https://unesdoc.unesco.org/ark:/48223/pf0000374866/PDF/374866eng.pdf.multi

Paulus, C. (2009). *Der Saarbruecker Persönlichkeitsfragebogen SPF (IRI) zur Messung von Empathie.* Universität des Saarlandes.

Pistone, I., Andersson, L. M., Lidström, A., Nelhans, G., Pernler, T., Sivenbring, J., Sager, M., & Mattson, C. (2024). Teaching and learning about the Holocaust: A systematic scoping review of existing research. *Holocaust Studies, 30*(1), 1–21. https://doi.org/10.1080/17504902.2023.2245282

Priebe, J. A. (2024). *Testkonstruktion – das Herz der psychologischen Diagnostik: Von Gütekriterien, lokaler stochastischer Unabhängigkeit, Normen und dem IQ.* Springer. https://doi.org/10.1007/978-3-662-67547-2

Renkl, A. (2014). Wissenserwerb. In E. Wild & J. Möller (Eds.), *Pädagogische Psychologie* (pp. 3–24). Springer.

Rogers, C. R. (2009). *Eine Theorie der Psychotherapie, der Persönlichkeit und der zwischenmenschlichen Beziehungen.* Reinhardt. (Original work published 1959)

Savenije, G. M., Brauch, N., & Wagner, W. (2019). Sensitivities in history teaching across Europe and Israel. *Pedagogy Culture and Society, 27*(1), 1–6. https://doi.org/10.1080/14681366.2019.1566163

Schweitzer, F., Bräuer, M., & Boschki, R. (Eds.). (2017). *Interreligiöses Lernen durch Perspektivenübernahme: Eine empirische Untersuchung religionsdidaktischer Ansätze.* Waxmann.

Sevicestelle für Antidiskriminierungsarbeit, Beratung bei Rassismus und Antisemitismus. (2025). *Malmad—Virtueller Methodenkoffer gegen Antisemitismus.* https://www.malmad.de/

Steedle, J. T., & Way, J. D. (2024). Decomposing high-school GPA as a predictor of first-year GPA. *Educational Assessment, 29*(1), 21–35. https://doi.org/10.1080/10627197.2024.2345599

Stefaniak, A., & Bilewicz, M. (2016). Contact with a multicultural past: A prejudice-reducing intervention. *International Journal of Intercultural Relations, 50*, 60–65. https://doi.org/10.1016/j.ijintrel.2015.11.004

Stefaniak, A., Wohl, M. J. A., Bilewicz, M., & Petelewicz, J. (2022). Leveraging knowledge about historical diversity: A meta-analysis of findings from the school of dialogue intergroup intervention. *Peace and Conflict: Journal of Peace Psychology, 28*(3), 314–326. https://doi.org/10.1037/pac0000626

Striewski, R. (2026, February 25). *Materialien gegen Judenhass: Portal soll Lehrkräfte im Kampf gegen Antisemitismus stärken.* WDR. https://www1.wdr.de/politik/politik-in-nrw/antisemitismus-schule-amudim-100.html

Thein, C., & Hiepel, L. (2024). *Konzeptpapier Antisemitismusprävention Lehrkräftebildung.* Universität Münster.

Tikvah Institut (Ed.). (in press). *Basiswissen Antisemitismusprävention: Ein Kerncurriculum für Staatsbedienstete.* De Gruyter.

Tourangeau, R., Rips, L. J., & Rasinski, K. A. (2000). *The psychology of survey response.* Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Trautwein, U., Bertram, C., Borries, B., Brauch, N., Hirsch, M., Klausmeier, K., Körber, A., Kühberger, C., Meyer-Hamme, J., Merkt, M., Neureiter, H., Schwan, S., Schreiber, W., Wagner, W., Waldis, M., Werner, M., Ziegler, B., & Zuckowski, A. (2017). *Kompetenzen historischen Denkens erfassen: Konzeption, Operationalisierung und Befunde des Projekts „Historical Thinking – Competencies in History" (HiTCH).* Waxmann.

Tribukait, M. (2021). Students' prejudice as a teaching challenge: How European history educators deal with controversial and sensitive issues in a climate of political polarization. *Theory & Research in Social Education, 49*(4), 540–569. https://doi.org/10.1080/00933104.2021.1947426

Vernikoff, L., Morvay, J. K., & Kolman, J. (2024). Where is antisemitism in teacher preparation? An exploration of the perspectives and practices of equity-oriented teacher educators. *Equity & Excellence in Education, 57*(2), 183–196. https://doi.org/10.1080/10665684.2022.2047414

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen—Eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (pp. 17–31). Beltz.

Wetzel, J. (2006). *Stichwort Antisemitismus* (Themenblätter im Unterricht, Nr. 56). Bundeszentrale für politische Bildung. https://www.bpb.de/shop/materialien/themenblaetter/36644/stichwort-antisemitismus/

Zachrich, L., & Grimm, M. (in press). Evidenzbasierung antisemitismuskritischer Bildung: Perspektiven empirischer Bildungsforschung und weiterführende Forschungsdesiderate. In R. Boschki, F. Schweitzer, & F. Ulfat (Eds.), *Braucht Antisemitismusprävention empirische Forschung?* (pp. 53–70). Waxmann.

Zentralrat der Juden in Deutschland, Kultusministerkonferenz, & Beauftragter der Bundesregierung für jüdisches Leben und den Kampf gegen Antisemitismus. (2025). *Handreichung der gemeinsamen Arbeitsgruppe Antisemitismus zur Verankerung des Themas Antisemitismus in der Lehrkräftebildung*. https://www.kmk.org/fileadmin/Dateien/pdf/Bildung/AllgBildung/2025-10-09_Handreichung-Antisemitismus-Lehrkraeftebildung.pdf

Zick, A., Küpper, B., Mokros, N., & Eden, M. (Eds.). (2025). *Die angespannte Mitte: Rechtsextreme und demokratiegefährdende Einstellungen in Deutschland 2024/25.* Dietz.

Appendix

1 Appendix A

Table A1: Item difficulty and discrimination of 2PL model for final item pool in test version 1 (Study 2).

Item No.	Keyword	*a*	*SE*	*b*	*SE*
K01	Shabbat	0	0.22	0.89	0.26
K02	Synagogue	2.61	1.38	0.43	0.23
K03	Plurality of Judaism 1	0.21	0.25	0.77	0.24
K04	Plurality of Judaism 2	0.13	0.24	0.79	0.25
K05	Plurality of Judaism 3	-0.54	0.26	0.87	0.27
K06	BDS campaign	0.3	0.17	1.5	0.37
K07	Antisemitism in Middle Ages	1.97	0.36	2.41	1.04
K08	Jewish deicide	-0.69	0.32	0.74	0.25
K09	Number of Jewish people in Germany 1933	1.56	0.31	1.61	0.47
K10	Fatalities of the Shoah	-0.92	0.22	1.54	0.45
K11	The establishment of the State of Israel	-0.81	0.24	1.2	0.35
K12	Jerusalem 1	1.18	0.74	0.4	0.21
K13	Jerusalem 2	1.19	0.33	1.03	0.30

K14	Red triangle	-0.4	0.30	0.67	0.24
K15	1948 Palestinian war	1.06	0.49	0.58	0.23
K16	Intifada	-0.32	0.19	1.18	0.32
K17	Zionism	-0.21	0.14	1.9	0.52

Note. 138 < *N* < 147; Keyword describes topic of the item, *a* = item difficulty, *b* = item discrimination

Table A2: Item difficulty and discrimination of 2PL model for final item pool in test version 2 (Study 2).

Item No.	Keyword	*a*	*SE*	*b*	*SE*
K18	Celebrations at the synagogue	2.52	0.75	0.97	0.36
K19	Torah shrine	-1.03	0.36	0.8	0.26
K20	Ethical principles	2.1	0.88	0.55	0.24
K21	Jewish people in Germany today	-0.11	0.33	0.53	0.21
K22	Head of Judaism	-0.43	0.18	1.31	0.33
K23	Secondary antisemitism	2.17	0.58	1.08	0.36
K24	Ritual murder libel	-0.48	0.17	1.62	0.40
K25	Development of antisemitism	-0.56	0.29	0.77	0.25
K26	The beginning of the Shoah	1.58	0.36	1.26	0.37
K27	Auschwitz-Birkenau	-1.98	0.62	0.95	0.35
K28	Night of Broken Glass	-1.98	0.47	1.73	0.68
K29	Operation Barbarossa	-0.65	0.28	0.86	0.26
K30	Jewish people in Germany 1933	1.66	0.60	0.67	0.25
K31	Chronicles of the Arab-Israeli conflict	0.53	0.18	1.49	0.37
K32	Map of the Middle East	0.54	0.39	0.53	0.22
K33	Diplomatic relationships GDR	0.06	0.25	0.78	0.25
K34	Israel’s democracy	0.24	0.23	0.9	0.27
K35	History of Jerusalem 1	0.48	0.15	1.95	0.50
K36	History of Jerusalem 2	0.17	0.19	1.14	0.30
K37	History of Jerusalem 3	0.67	0.20	1.35	0.35
K38	Guarantor’s obligation teachers	-0.11	0.37	0.5	0.22

Note. 135 < *N* < 148; Keyword describes topic of the item, *a* = item difficulty, *b* = item discrimination

Table A3: Item difficulty and discrimination of 2PL model for final item pool in test version 3 (Study 2).

Item No.	Keyword	*a*	*SE*	*b*	*SE*
K39	Jewish ethnic divisions	0.40	0.28	0.74	0.26
K40	German attack on the USSR	0.77	0.42	0.61	0.26
K41	Murdered Jews in Germany	2.36	0.67	1.43	0.62
K42	Largest Jewish community before 1939	-0.18	0.38	0.51	0.23
K43	Country with largest death toll	-0.15	0.34	0.56	0.25
K44	National Socialism in power	-1.72	0.41	1.55	0.56
K45	Definition of Zionism	-0.84	0.34	0.9	0.32
K46	Jewish presence in the Middle East	-0.41	0.29	0.78	0.28
K47	Being Jewish	1.27	0.34	1.16	0.37
K48	“Great Replacement”	0.46	0.35	0.60	0.24
K49	Functions of antisemitism	-0.36	0.21	1.11	0.34
K50	Antisemitism as hatred against…	2.02	0.92	0.57	0.27
K51	“Modern antisemitism” 1	1.34	0.28	1.78	0.59
K52	“Modern antisemitism” 2	-0.47	0.27	0.87	0.29
K53	Countries with highest death tolls	1.51	0.41	1.16	0.40
K54	Camp David Accords	0.03	0.22	0.98	0.31
K55	Religious symbols	-0.96	0.24	1.52	0.46

Note. 125 < *N* < 138; Keyword describes topic of the item, *a* = item difficulty, *b* = item discrimination.

2 Appendix B

Table B: Descriptive statistics of competence items and correlation with WLE score in Study 2.

Item No.	Keyword	*N*	*M*	*SD*	*r* WLE
C1	“Israel, child murderer”	134	0.24	0.43	**.30**
C2	Social media scenario	134	0.29	0.46	.03
C3	“From Berlin to Gaza—Yallah intifada”	134	0.70	0.46	**.36**
C4	“Al-Quds Day”	134	0.60	0.49	**.35**
C5	Antisemitic motivation	134	0.61	0.49	-.05
C6	“Free Palestine from German guilt”	135	0.37	0.48	.14
C7	“Israel, child murderer” 2	123	0.24	0.43	.09

C8	“From Lützi to Gaza—Climate intifada”	123	0.56	0.50	.02
C9	Arab-Israeli conflict in a classroom scenario	123	0.43	0.50	.08
C10	“You Jew” in a classroom scenario	123	0.43	0.50	-.11

Note. Statistically significant correlations with WLE knowledge on .05-level are printed in bold.

3 Appendix C

Table C: Item statistics, item difficulty and discrimination for each item (2PL model, Study 3) for knowledge items in Study 3.

Item # Study 3	Item # Study 2	Keyword	*M*	*SD*	r_{is}	*a*	*SE*	*b*	*SE*
K01	K19	Torah shrine	0.20	0.4	0.29	1.85	0.1	0.86	0.06
K02	K01	Shabbat	0.36	0.48	0.33	0.77	0.1	0.86	0.06
K03	K22	Head of Judaism	0.47	0.5	0.32	0.16	0.1	0.85	0.06
K04	K47	Being Jewish	0.15	0.36	0.36	1.87	0.1	1.15	0.08
K05	-	Immigration to Germany	0.32	0.46	0.30	1.14	0.1	0.76	0.06
K06	K02	Synagogue	0.15	0.35	0.26	2.43	0.2	0.81	0.07
K07	K20	Ethical principles	0.19	0.39	0.28	2.09	0.1	0.79	0.06
K08	K25	Development of antisemitism	0.4	0.49	0.38	0.48	0.1	1.03	0.06
K09	K48	“Great Replacement”	0.07	0.26	0.29	1.63	0.1	1.08	0.07
K10	K49	Functions of antisemitism	0.38	0.49	0.39	0.54	0.0	1.12	0.07
K11	K09	Jewish people in Germany 1933	0.12	0.33	0.32	2.23	0.1	1.06	0.08
K12	K41	Murdered Jews in Germany	0.1	0.3	0.21	3.46	0.3	0.7	0.08
K13	-	Two-state solution	0.67	0.47	0.28	-1.01	0.1	0.82	0.06
K14	K45	Definition of Zionism	0.38	0.49	0.39	0.58	0.1	1.03	0.06
K15	K14	Red triangle	0.41	0.49	0.21	0.77	0.1	0.52	0.05
K16	K34	Israel’s democracy	0.2	0.4	0.48	1.2	0.1	1.71	0.10
K17	K46	Jewish presence in the Middle East	0.36	0.48	0.35	0.71	0.1	0.93	0.06
K18	K15	1948 Palestinian war	0.26	0.44	0.26	1.69	0.1	0.67	0.06
K19	K31	Chronicles of the Arab-Israeli conflict	0.16	0.36	0.42	1.56	0.1	1.47	0.09

K20	K38	Guarantor's obligation teachers	0.35	0.48	0.31	0.84	0.1	0.83	0.06
K21	-	Criminal acts	0.01	0.08	-0.03				
K22	K50	Antisemitism as hatred against...	0.07	0.25	0.20	2.44	0.2	0.63	0.06
K23	K52	"Modern antisemitism" 2	0.49	0.5	0.31	0.05	0.1	0.82	0.06
K24	K23	Secondary antisemitism	0.12	0.32	0.32	2.19	0.1	1.13	0.08
K25	-	Statements about antisemitism	0.46	0.5	0.27	0.24	0.1	0.7	0.05
K26	-	Right-extremist codes	0.03	0.16	0.18	2.82	0.2	1.1	0.09
K27	K27	Auschwitz-Birkenau	0.72	0.45	0.31	-1.1	0.1	1.04	0.07
K28	K10	Fatalities of the Shoah	0.65	0.48	0.29	-0.83	0.1	0.83	0.06
K29	K29	Operation Barbarossa	0.46	0.5	0.32	0.21	0.1	0.8	0.06
K30	K42	Largest Jewish community before 1939	0.44	0.5	0.16	0.72	0.1	0.38	0.05

Note. 2,555 < *N* < 2,991. Cronbach's Alpha α = .78, a = item difficuty, b = item discrimination, WLE PSR = 0.66. In final 2PL model, K21 was excluded.

4 Appendix D

Table D: Descriptive statistics and correlations for competence items in Study 3.

Item # Study 3	Item # Study 2	Keyword	*M*	*SD*	C4	C6	C7
C1	C3	"From Berlin to Gaza-Yallah Intifada"	0.09	0.29	.12	.14	.02
C2	C4	"Al-Quds Day"	0.40	0.49		.20	.04
C3	C6	"Free Palestine from German Guilt"	0.17	0.38			.07
C4	C7	"Israel, Child Murderer"	0.17	0.37			

Note. 2,556 < *N* < 2,596; Items were scored with 1 = *every answer correct*, 0 = *at least one an-swer incorrect.*