Journal of Public Health
https://doi.org/10.1007/s10389-026-02683-6

ORIGINAL ARTICLE

Developing a measurement instrument for physical activity-related health literacy (PA-HL): results from an eDelphi and cognitive interviews

Elena Fitzner[1,2] · **Thomas Hering**[3] · **Kevin Dadaczynski**[1,4]

Received: 24 September 2025 / Accepted: 13 January 2026

Abstract

Aim Using the health literacy (HL) framework proposed by Sørensen et al. (BMC Public Health 12:80, 2012), this study aimed to develop a tool for measuring information processing competence focused on physical activity (PA). This paper presents results from an approach using the online Delphi (eDelphi) method along with cognitive interviews for the development of a comprehensive physical activity-related HL measurement instrument (PA-HL).

Subject and methods PA is a major determinant in preventing noncommunicable diseases and promoting health. Access to and appropriate use of evidence-based PA information is important for making informed decisions favoring PA, particularly in cases of an overabundance of information. Here, a two-stage development process was applied: (1) A three-round eDelphi strategy was employed with HL and PA experts, where items were rated according to relevance and comprehensibility. Consensus was operationalized as an item content validity index; the second round incorporated an online group workshop. (2) Cognitive interviews were conducted with adult participants, utilizing think-aloud and probing techniques. The item pool was analyzed utilizing a questionnaire appraisal coding scheme.

Results The preliminary measurement instrument consisted of 60 items. After revision through both stages, it was reduced to 33 items. Most changes addressed undefined terms and overly complex questions. Cognitive interviews suggested the addition of five items and clearly distinguishing between "apply" and "decide" competencies to ensure conceptual clarity.

Conclusion The eDelphi study confirmed high content validity, while cognitive interviews demonstrated that most items were well understood but required linguistic simplification and shortening of lengthy or complex items. The item pool strongly focused on information processing competencies which involve decision-making, rather than behavior itself. Future research will psychometrically evaluate the PA-HL measurement instrument.

Keywords Delphi study · Cognitive interviews · Physical activity · Health literacy · Measurement instrument

Background

From a public health perspective, physical activity (PA) is a key determinant of health, contributing to the prevention and management of noncommunicable diseases (NCDs) and the promotion of well-being, addressing all three areas of health literacy (disease management, disease prevention, and health promotion) (Buecker et al. 2021; Garcia et al. 2023; Warburton and Bredin 2017; World Health Organization 2024). Consequently, strategies for reducing levels of physical inactivity represent an important priority of public health policies (Bundesvereinigung Prävention und Gesundheitsförderung e.V 2025; World Health Organization 2024). Findings from a meta-analysis show significant associations between PA levels and mortality reduction, alongside decreased cardiovascular disease and incidence

✉ Elena Fitzner
elena.fitzner@uni-potsdam.de

1 Faculty of Human Science, Department Sports and Health Sciences, University of Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany

2 Westphalian University of Applied Sciences - Institute for Work and Technology, Research Department Health industries & Quality of Life, Munscheidstraße 14, 45886 Gelsenkirchen, Germany

3 Department of Applied Human Sciences, University of Applied Sciences Magdeburg-Stendal, Osterburger Straße 25, 39576 Stendal, Germany

4 Centre for Applied Health Science, Leuphana University Lueneburg, Wilschenbrucher Weg 84a, Lüneburg, Germany

Published online: 21 February 2026

of cancer (Garcia et al. 2023). Furthermore, it is estimated that potentially 15.7% (95% CI 13.1–18.2) of all premature mortality could be averted if the individuals who were insufficiently active had an adequate level of PA (8.75 marginal metabolic equivalent of task (mMET)-hours/week) (Garcia et al. 2023). In Germany, sedentary behavior is projected to account for approximately 6% of total cancer cases between 2024 and 2040, ranging between 2% and 9% depending on specific cancer types and sex, which results in a substantial economic burden (Schaber et al. 2025). Despite such well-documented research evidence, the global age-standardized prevalence of insufficient physical activity has progressively increased, from 23.4% (95% uncertainty interval: 21.1–26.0) in 2000, to 26.4% (24.8–27.9) in 2010 to 31.3% in 2022 (28.6–34.0) (Strain et al. 2024). According to the DKV (Deutsche Krankenversicherung) Report of 2025, more than two-thirds of the German population engage in sufficient physical activity to achieve health benefits. Nevertheless, the findings indicate that approximately one-fifth of Germans are classified as physically inactive, which constitutes a significant risk to their health (Froböse and Wallmann-Sperlich 2025).

At the same time, the information environment surrounding PA has become increasingly complex, presenting significant challenges in information acquisition, comprehension, assessment, and use in everyday life. Information from web-based PA reveals substantial inconsistencies in messaging content, with inadequate translation of knowledge from evidence-based PA guidelines into information resources for everyday use (Thomas and Cardinal 2020). Research from 2009 revealed that only 22% of online PA information sources fulfilled standards of high quality, with none achieving high accuracy. These findings suggest that individuals seeking guidance for PA face potential harm due to low-quality or unreliable information (Bonnar-Kidd et al. 2009). Additionally, population-level interventions such as mass media information campaigns promoting PA have demonstrated limited efficacy (Abioye et al. 2013). For example, while such campaigns may encourage an increase in walking behavior, evidence indicates that they fail to significantly reduce sedentary behavior or increase overall PA to recommended levels (Abioye et al. 2013).

The dissemination of information regarding PA, among other topics, increasingly takes place through social media channels, facilitating self-management of PA behavior through the provision of training, alongside opportunities for peer, familial, and social group interaction (Goodyear et al. 2021). Various information providers, including social media influencers, interpersonal networks, and official organizations, significantly influence which PA-related information is accessed via social media platforms (Goodyear et al. 2021). Furthermore, the current information ecosystem lacks a comprehensive quality assessment framework for content from fitness influencers. Kaňková et al. (2024) reported that health experts generally regard the dissemination of information via social media as beneficial, while simultaneously emphasizing the associated risks. These health experts observed that in the social media world, even nonexperts position themselves as experts, basing their advice on personal experiences. This can result in the sharing of incorrect information or overgeneralization, which may be harmful. In terms of exercise, the study highlights that generalized information may not be appropriate in specific contexts: certain types of exercise can be highly beneficial for particular groups but potentially harmful for others, or should only be undertaken following consultation with a qualified professional (Kaňková et al. 2024). These findings highlight the need for adequate information-processing competencies, which include the ability to critically appraise information.

In infodemic times, the ability to deal with PA-related information in a meaningful and appropriate manner is becoming increasingly important. Therefore, it is crucial for individuals to be able to find, understand, and appraise PA-related information, and to apply it appropriately to their own health situation or to that of others.

Because the proliferation of health information requires effective navigation, health literacy (HL) has garnered substantial attention from both researchers and practitioners. General HL levels remain inadequate in the German population and worldwide (Schaeffer et al. 2021). Addressing this shortfall is important; empirical research has consistently demonstrated associations between health literacy and health behavior (Buja et al. 2020; Goto et al. 2018; Xie et al. 2022), and it is assumed that enhanced HL skills can lead to improved health outcomes throughout a person's life course (Sørensen et al. 2012, 2013). Conceptual heterogeneity within HL research has led to the creation of diverse definitional frameworks and measurement instruments for epidemiological and interventional purposes (Griebler et al. 2025; Wirtz and Soellner 2022). One common definition, which resulted from a literature review by Sørensen et al. (2013), defines HL as "people's knowledge, motivation and competences to access, understand, appraise and apply health information in order to make judgments and take decisions in everyday life concerning health care, disease prevention and health promotion to maintain or improve quality of life during the life course" (p. 3). Based on this definition, the 12-item version of the European Health Literacy Survey Questionnaire (HLS-Q12) was developed and is among the most widely used research instruments for general HL assessment. It consists of one item from each of 12 dimensions (Griebler et al. 2025). Over the years, a wide range of HL measurement instruments have been developed, with a trend toward instruments for assessing HL in specific areas, for example, disease- or lifestyle-related factors (Griebler et al. 2025) such as navigational HL (Griese and Schaeffer 2022), communicative HL (Finbråten et al. 2022), movement-related HL (Sudeck et al. 2022), HL within the context of the COVID-19 pandemic (Matterne et al. 2022), or in a dental setting (Schladitz 2019).

Currently, no measurement instrument is available that specifically targets PA-related health literacy (PA-HL) with a focus on information processing competencies, as conceptualized by Sørensen et al. (2012, 2013). Accordingly, based on empirical findings, the objective of this paper is to develop and validate a measurement instrument for PA-HL that focuses on the processing of PA-related health information. The PA-related information processing competencies present a perspective as a potential determinant for PA. The results of the development of a comprehensive PA-HL measurement instrument using an online Delphi (eDelphi) technique and cognitive interviews are presented. The following research questions guided the process:

1. How do experts judge the item pool of the PA-HL measurement instrument in terms of relevance and comprehensibility, and what recommendations do they provide for refinement?
2. How do adults as prospective users understand and assess the items of the PA-HL measurement instrument?

Methods

The development process for the PA-HL measurement instrument is based on the guidelines of Boateng et al. (2018), which define scale development and validation procedures for health, social, and behavioral research, using a three-stage process. The first step is the item development, encompassing the identification of domains and item generation, as well as aspects of content validity. The second step is scale development, incorporating for example pretesting and item reduction. The third step focuses on scale evaluation, involving tests of dimensionality, reliability, and validity. The rationale and study design of the PA-HL development and validation are described elsewhere (Fitzner et al. 2025). This manuscript specifically addresses the first and second procedural steps from the guidelines of Boateng et al. (2018): the implementation of an eDelphi process and cognitive interviews.

Delphi procedures are structured group communication processes in which complex issues, where knowledge is uncertain or incomplete, are assessed by experts through an iterative process (Niederberger and Renn 2019). Delphi techniques are utilized to systematically elicit expert opinions and achieve consensus (Köck-Hódi and Mayer 2013) and are a significant source of evidence in healthcare research (Jünger et al. 2017). The eDelphi technique assembles ideas online via experts communicating and engaging with one another in their own time: a systematic and iterative consensus-building methodology for research and development issues, such as the relevance or comprehensibility of items (Niederberger and Renn 2019). Delphi techniques have been demonstrated as an innovative, systematic, and reliable approach (Häder 2009; Niederberger and Renn 2019). Criteria for establishing trust in Delphi techniques can comprise such qualities as confirmability, credibility, dependability, and transferability (Niederberger and Renn 2019).

In addition, cognitive interviews were conducted in order to refine the instrument and its validation within the target population (18+ years), assess the appropriateness of the item pool relative to measurement objectives (Bethmann et al. 2019; Knafl et al. 2007; Lenzner et al. 2024; Pohontsch and Meyer 2015), and identify potential limitations in terms of item interpretation and response formulation (Brancato et al. 2006; Pohontsch and Meyer 2015). Ethical approval was granted by the independent ethics committee of Fulda University of Applied Sciences on December 21, 2023, and June 21, 2024 (reference EK231204).

Study design and methods

eDelphi

The eDelphi process was pretested (Köck-Hódi and Mayer 2013) with eight independent experts in January 2024 to evaluate questionnaire functionality, comprehensibility parameters, and logical sequence integrity. These experts did not participate in the resulting eDelphi study. Experts participating in the eDelphi study were provided with a brief introduction to the study's topic and objectives, the definitions of PA and HL, and details on the procedure. The eDelphi was structured in three iterative rounds. Subsequent round invitations were distributed only to participants who completed the preceding rounds. After each iteration, the item pool underwent systematic modification based on the outcomes (see statistical analysis). At the beginning of a new round, the anonymized results were sent to all participants.

Round 1

The initial eDelphi round was conducted as an online survey utilizing the Unipark platform (Tivian EFS v.24.1), with data collection taking place over a minimum period of 6 weeks (March and April 2024). Participants were instructed to evaluate individual items in terms of both relevance and comprehensibility and to propose potential areas for refinement. A four-point Likert scale was used to assess relevance (very relevant, relevant, moderately relevant, irrelevant) and comprehensibility (very easy, easy, difficult, very difficult).

Round 2

The second eDelphi round incorporated an online workshop (April 2024) consistent with a group Delphi or expert group

protocol (Brancato et al. 2006). Prior to active participation, participants were presented with an informed consent form through which they would subsequently provide their consent. The synchronous interactive workshop format facilitated the acquisition of extensive participant feedback and enabled in-depth discussions of underlying rationales (Niederberger and Renn 2019). This group Delphi approach represents a mixed-method integration of qualitative elements (expert narrative contributions). Discussion topics were derived from the outcomes of round 1, including the fundamental structure and design of the measurement instrument, and items that were given contradictory ratings on relevance and comprehensibility.

Round 3

The third round employed the same online survey procedure as in round 1, and included 2 months of data collection (May and June 2024). Round 3 aimed to systematically evaluate the modified item pool derived from workshop proceedings and elicit feedback regarding item relevance, comprehensibility, and potential refinements, utilizing the same four-point Likert-type scales.

Cognitive interviews

Upon completion of the eDelphi process, the revised PA-HL measurement instrument was presented to a target group via cognitive (online) interviews (August to November 2024). The primary objective was to assess the underlying cognitive mechanisms in item processing (e.g., comprehension, information retrieval, judgment, and response behavior; Tourangeau et al. 2000). A semi-structured methodological approach incorporated standardized guidelines with general, specific, and memory probes supplemented by spontaneous interrogatives (Brancato et al. 2006; Lenzner et al. 2024). The cognitive interviews started with open-ended questions addressing conceptual domains of PA and HL. Through the think-aloud technique, participants were asked to verbalize their thoughts when answering the questions on each item. Probes were used to evaluate question comprehension and response quality (Ericsson and Simon 1980; Pohontsch and Meyer 2015). Representative probe questions are presented in Table 1. In addition to feedback on the items, participants were also asked to evaluate the appropriateness of response categories. At the end, participants completed a questionnaire on sociodemographic characteristics, their health literacy level (HLS-Q12 score), and physical activity (International Physical Activity Questionnaire–Short Form [IPAQ-SF]).

Participants

The selection of experts for the eDelphi study was guided by searching for candidates with scientific expertise or experience in the fields of HL, sport science, physiotherapy, and public health, with respective activities or publications serving as inclusion criteria. After an online search, a list of potential experts was systematically compiled, and selected experts were then contacted via email invitation.

A total of 50 experts were initially invited to participate in the eDelphi study. Nineteen experts participated in the first eDelphi round, 12 in the second round, and 11 in the third round. Despite systematic reminders and repeated contact, participant retention decreased after each round. Female participants were somewhat overrepresented, and the mean age ranged from 40 to 42 years. The predominant professional backgrounds were professors or research assistants.

A convenience sample of adults aged 18 years and older was recruited through a snowball approach for the cognitive interviews. A participant sample size of between 5 and 15–30 individuals is recommended for cognitive interviews (Lenzner et al. 2024; Prüfer and Rexroth 2005). Demographic variables (e.g., gender, age, education, migration) (Beatty and Willis 2007) were taken into account when selecting from the participants pool. Accordingly, the research project was presented in various professional and personal contexts and networks in order to solicit potential participants. In keeping with the snowball approach, interview participants were asked to suggest additional potential candidates.

Table 1 Examples of cognitive interview probe questions used to assess survey item comprehension and response processes

Category	Probes
Understanding and interpretation	• What kind of physical activity are you thinking of? • Can you explain to me how you got to this answer?
Paraphrase	• Can you please repeat the question in your own words? • Could you tell me, in your own words, what you thought this question was asking?
Reminder probes	• What recommendations for physical activity do you know?
General probes	• What do you think of when you read the words “disease prevention”? • And which professionals come to mind?
Specific probes	• This question should make you think of information about physical activity. • What do you understand by trustworthy information? • What did you think was meant by over- or underload?

A total of 19 adults participated in the cognitive interviews, with equal gender distribution achieved and a mean age of 44 years. Participants with a higher level of education were somewhat overrepresented in the study group, and 79% reported no migration background. The general HL of participants (based on the HLS-Q12 scoring) could be classed as "problematic" (42%), while participants were physically very active (based on the IPAQ) (58%). A summary of the characteristics of participants in both studies is provided in Table 2.

Data analysis

Statistical analysis was performed using IBM SPSS Statistics, v.24.0 (IBM) for quantitative data analysis. Descriptive statistical parameters were calculated for sociodemographic variables.

Quantitative data analysis

Descriptive statistics were computed for the eDelphi rounds 1 and 3, including measures of central tendency (mean) and dispersion (standard deviation) (Köck-Hódi and Mayer 2013). An item content validity index (I-CVI) was calculated (Polit et al. 2007) to determine whether items were to be included or excluded. A four-point Likert scale was used to assess both relevance (1 = very relevant, 2 = relevant, 3 = moderately relevant, 4 = irrelevant) and comprehensibility (1 = very simple, 2 = simple, 3 = difficult, 4 = very difficult). The rating scale for relevance was dichotomized and transformed into a binary variable whereupon raw values 3 and 4 were recoded as "0" (not relevant) and raw values 1 and 2 as "1" (relevant), thereby generating two discrete response categories: "relevant" and "not relevant" (Lynn 1986). The following formula for transformation was applied: f: X -> Y, with f(x) = [1 for x ∈ {1,2}, 0 for x ∈ {3,4}]. Composite expert ratings of the transformed items (0 and 1) were summed and divided by the total number of expert evaluators, resulting in generated values ranging from 0 (no expert classified an item as relevant) to 1 (unanimous expert classification of an item as relevant). The value for exclusion was an I-CVI below 0.78 (Lynn 1986). In this study, the I-CVI was interpreted in conjunction with qualitative feedback data. Furthermore, an average scale content validity index (S-CVI/Ave) was calculated in round 3 by dividing the sum of all I-CVI values by the total number of items (Polit et al. 2007). An S-CVI/Ave threshold of 0.90 was then used to accept the scale (Waltz et al. 2010).

Qualitative data analysis

The reporting of qualitative data adheres to the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist (Tong et al. 2007). The eDelphi workshop was designed to facilitate open discussion on opinions and ratings from the first eDelphi round, and to determine whether diverging judgments could be resolved by dialogue (Niederberger and Renn 2019). Workshop proceedings were audio-recorded and transcribed (Mayring and Fenzl 2019), and the transcripts examined through content analysis (Niederberger and Renn 2019). The general structure of the measurement instrument and the individual items were then adapted based on the discussion results (Niederberger and Renn 2019).

The cognitive interviews were digitally recorded, transcribed verbatim, and formatted into a standardized data analysis sheet. Subsequently, item-specific files were generated containing transcriptions of each respondent's interpretation of that item along with supplementary comments and probe responses (Knafl et al. 2007; Prüfer and Rexroth 2005). Analysis was performed on the questions to identify potential response bias (Knafl et al. 2007; Prüfer and Rexroth 2005). Response data were analyzed using the deductive categorical framework of the questionnaire appraisal coding scheme (QACS) (DeMaio and Landreth 2004). The QACS represents a comprehensive analytical framework for the systematic identification and categorization of the potential limitations of a measurement instrument. This validated scheme operationalizes four primary domains of assessment: (1) comprehension difficulties, encompassing question content and structural ambiguities; (2) retrieval challenges, addressing memory access and temporal reference limitations; (3) judgment complexities, involving estimation processes and potential response biases; and (4) response selection inhibitions, including terminological imprecision and categorical inadequacies (DeMaio and Landreth 2004). Based on the QACS categorical classifications and further empirical research findings on construct differentiation (Griebler et al. 2025), the item pool was linguistically refined, expanded, or reduced in regard to information processing competencies. Analytical uncertainties were resolved through a deliberative consensus by a multidisciplinary research team comprising five researchers in order to facilitate appropriate item adaptation.

Results

The development of an initial item pool for the PA-HL measurement instrument was based on existing HL instruments, including the 47-item European Health Literacy Survey Questionnaire (Sørensen et al. 2012, 2013), Health Literacy Instrument for Adults (Tavousi et al. 2020), All Aspects of Health Literacy Scale (Chinn and McCarthy 2013), and physical activity-related health competence (PAHCO) model (Carl et al. 2024), and systematic reviews on predictors of PA (Benda 2019; Rhodes and Quinlan 2015) and research on PA information-seeking behavior (Idowu et al. 2024; Lee et al. 2022), resulting in

Table 2 Sociodemographic characteristics of participants in the eDelphi process (n = 19–11) and cognitive interviews (*n* = 19), as part of the development of a physical activity-related health literacy measurement instrument, 2024

	eDelphi			**Cognitive interviews**	
	Round 1 *(online survey)*	**Round 2** *(online workshop)*	**Round 3** *(online survey)*		
Information about the eDelphi rounds				**Information about the interviews**	
	M (SD) [95% CI]				M (SD) [95% CI]
Mean duration (in min)	58:61 (30.00) [43.91; 73.30]	75	27.21 (14.38) [17.82; 36.61]	Mean duration (in min)	58:35 (17:45) [50:36; 66:34]
Total no.	19	12	11	Total no. (%)	19 (100)
Response rate	38%	63%	92%		
Sociodemographic characteristics					
Gender	No. (%)			**Gender**	No. (%)
Female	15 (78.95)	8 (66.67)	7 (63.64)	Female	9 (47.37)
Male	4 (21.05)	4 (33.33)	4 (36.36)	Male	10 (52.63)
	M (SD) [95% CI]			**Age** (years)	M (SD) [95% CI]
Age (years)	41.63 (11.53) [36.45; 46.82]	40.08 (11.12) [33.79; 46.38]	39.6 (11.89)		43.58 (14.29) [37.15; 50.01]
Professional title	No. (%)			**Education**	No. (%)
Professor	5 (26.32)	3 (25)	2 (18.18)	Compulsory basic secondary schooling	4 (21.05)
Postdoctoral researcher	3 (15.79)	1 (8.33)	3 (27.27)	Secondary education	3 (15.79)
Research assistant	6 (31.58)	7 (58.33)	3 (27.27)	General university entrance qualification	7 (36.84)
Physician	4 (21.05)	0 (0)	3 (27.27)	Bachelor's degree	0 (0)
Other	1 (5.26)	1 (8.33)	0 (0)	Master's degree	4 (26.67)
Work experience	No. (%)			Doctorate	1 (5.26)
Under 1 year	0 (0)	0 (0)	0 (0)		
1–3 years	1 (5.26)	1 (8.33)	1 (9.09)	**Migration background**	No. (%)
3–5 years	2 (10.53)	1 (8.33)	1 (9.09)	Yes	4 (21.05)
5+ years	3 (15.79)	4 (33.33)	3 (27.27)	No	15 (78.95)
				General HL	M (SD) [95% CI]/Mdn (IQR) [range]
				HLS-Q12 score	57.90 (23.18) [47.47; 68.32]/58.33 (50.00; 75.00) [0; 91.67]
					No. (%)
				Excellent	4 (21.05)
				Sufficient	2 (10.53)
				Problematic	8 (42.11)
				Inadequate	5 (26.32)
				IPAQ-SF	No. (%)
				Inactive/low	0 (0)
				Minimally active/ moderate	8 (42.11)
				HEPA active/high	11 (57.89)

M mean, *SD* standard deviation, *95% CI* 95% confidence interval, *Mdn* median, *IQR* interquartile range, *min* minutes, *HEPA* health-enhancing physical activity

a first version with 60 items. Existing HL instruments were used to develop an initial structure for the PA-HL measurement instrument and to guide wording decisions, but primarily to enable content-related delimitation of the items, to ensure clear differentiation between existing measurement instruments and the PA-HL, and to avoid content overlap.

eDelphi

The feedback from all three eDelphi rounds is presented in the following sections, with the caveat that there was some contradictory expert feedback over the three rounds.

Round 1

In the first round, 23 items were excluded due to an I-CVI below 0.78 and another five items due to content feedback (content repetition/redundancy). In total, of the items excluded after round 1, eight items were from the healthcare dimension, 12 were from disease prevention, and eight were from health promotion, resulting in a fairly even distribution across all three dimensions. Most items were sufficiently comprehensible to the participants, but due to the qualitative feedback in which a heterogeneous target group was to be considered, all items were revised in order to simplify the sentence structure. For instance, experts suggested that certain terminology (e.g., relating to treatment, therapy, or treatment options) should be standardized, or that in certain cases, rather than referring to specialist experts or specialist personnel, specific examples should be provided. It was also suggested that professional terms (such as prevention) be avoided and replaced with more readily comprehensible descriptions (e.g., avoidance of disease). Only one item remained linguistically unchanged, while all other included items were linguistically standardized or revised. Explicit feedback from two experts indicated the need for a definition of PA, since the concept of PA is not self-explanatory and because a conceptual clarification of PA is relevant for all items. Accordingly, a brief definition of PA was included at the beginning of the PA-HL measurement instrument. Moreover, some experts recommended a harmonization of similar terms, and some suggested combining the disease prevention and health promotion domains. However, the act of combining the two dimensions was not undertaken due to the pre-existing foundational theoretical structure of HL by Sørensen et al. (2012, 2013), even though it was discussed with the experts in the workshop in the second round.

Round 2

In contrast to the recommendation from round 1 to provide a definition of PA at the beginning of the PA-HL measurement instrument, the consensus from round 2 was to avoid a definition of PA and instead to include precise examples of PA for each of the items, since nonexperts would most likely not understand the use of a general term or a comprehensive meaning for PA. Consequently, a PA example was incorporated into each item: "(e.g., through domestic work/gardening or walking,...)," or "(e.g., hiking or stair climbing,...)," or "(e.g., walking or swimming,...)." The idea to combine the two dimensions of disease prevention and health promotion and the underlying HL model from Sørensen et al. (2012, 2013) with its 3×4 matrix was discussed, and the consensus was that the two dimensions should not be combined. However, it was also agreed that differentiating between disease prevention and health promotion is difficult for laypersons, and thus the items should be formulated in such a way that participants do not feel that these dimensions overlap with one another. Based on the discussion, the wording of the items in the disease prevention and health promotion dimensions was revised: In disease prevention, the focus was placed more on preventing illness, whereas in health promotion, the focus was more on staying healthy or on healthy aging.

Round 3

A total of 32 items were initially presented for the online survey of round 3. After the statistical analysis of the I-CVI and feedback from the qualitative responses, 29 items were included in the PA-HL measurement instrument which was used for the cognitive interviews. Of the exclusions, one item was within the disease prevention dimension, and two within the health promotion dimension. This resulted in a remaining ten items in the healthcare dimension, ten items in the disease prevention dimension, and nine items in the health promotion dimension. It was noted by the experts that the recommendation to insert examples of PA as part of the item description disrupted the reading flow, who instead recommended a description of PA at the beginning (as suggested in round 1). These recommendations were implemented for the cognitive interviews: The examples of PA were omitted, and a brief definition of PA was included at the beginning of the measurement instrument: "Physical activity is understood to encompass not only sport, but many types of movement, such as: cycling, walking, gardening, swimming, climbing stairs, and many others." The S-CVI/Ave of the entire scale is rated as acceptable (0.90). A summary of the analytical process and results of the eDelphi study are shown in Table 3.

Cognitive interviews

In accordance with the recommendations for sample size (Lenzner et al. 2024; Prüfer and Rexroth 2005), 19 participants (Table 1) were recruited for cognitive interviews in order to

Table 3 Items included and excluded by the eDelphi rounds, based on item content validity index thresholds and overall scale-level content validity index, 2024

	I-CVI	No. of items												
		Healthcare				Disease prevention				Health promotion				**Total**
		Fin	Und	Apr	Apl	Fin	Und	Apr	Apl	Fin	Und	Apr	Apl	
Round 1 (n = 19)	Included: 0.89–1.00	3	2	2	3	5 (D)	2	3	2 (D)	2	2	2 (D)	4	32
	Excluded: 0.52- 0.78	3	1	3	1	1	4	3	4	5	2	1	0	28
Round 2 (n = 12)	Included: *(no I-CVI)*	3	2	2	3	4	2	3	2	2	2	3	4	32
	Excluded: *(no I-CVI)*	0	0	0	0	0	0	0	0	0	0	0	0	0
Round 3 (n = 11)	Included: 0.89–1.00	3	2	2	3	3	2	3	2	2	2	3	2	29
	Excluded: 0.55–0.78	0	0	0	0	1	0	0	0	0	0	0	2	3
S-CVI/Ave	0.90													

Fin find information, *Und* understand information, *Apr* appraise information, *Apl* apply information, *I-CVI* item content validity index, *S-CVI* score content validity index (round 3), *(D)* indicates areas that had to be discussed in round 2, n = number of participating experts

assess the clarity and comprehensibility of the items. Analysis of the interview data revealed five important themes: (1) general comprehensibility, (2) unfamiliar terms or ambiguous interpretations (e.g., QACS C4, C7, R18), (3) challenging levels of abstraction (e.g., QACS J20), (4) missing content, and (5) response options.

General comprehensibility

Participants defined PA broadly, that is, beyond merely sport activities, and demonstrated a clear understanding of the definition of PA in the PA-HL measurement instrument. Although the participants did not suggest any improvements to the introductory text, analysis revealed a need to further clarify the term "information." A brief explanation now clarifies that the term "information" covers a variety of types and sources. The comprehensibility of the items was predominantly positive; most items were answered immediately or rated as easy to understand, and no uncertainties arose during the response process (58.44% of response options were positive). Five items were adapted slightly in terms of language, through minor clarifications (e.g., long car journeys), improved sentence structure, or adapted examples (e.g., swimming pools). Further information on the data analysis of each item in the cognitive interviews can be found in the additional file (Table 5), and the comparison of the item pools before and after the eDelphi procedure (prior to psychometric evaluation) (Table 6).

Unfamiliar terms or ambiguous interpretations (e.g., QACS Nos. 4, 7, 18)

For the remaining 41.56% of response options, frequencies and percentages for the QACS categories were calculated (Table 4). The most frequent QACS category, with more than a quarter (28.38%) of total responses, was "Question content, No. 7: undefined/vague term."

For some items, the participants perceived the examples of information sources (e.g., "to assess whether information (e.g., in magazines, adverts, on the internet,...) […] is trustworthy") as unsuitable; magazines and adverts evoked negative associations as merely promoting product sales. It was suggested to include different examples, and these examples were removed from the items. Instead, specific examples were included as part of the introductory text (e.g., "conversations with healthcare professionals, (online) articles, videos, podcasts, informational brochures, or data from fitness trackers and smartphones").

In addition, the word "recommendations" (e.g., "to understand recommendations regarding the type, duration and frequency of PA in order to prevent the occurrence of diseases") often caused irritation, and questions were repeatedly raised as to what these recommendations were or who they came from. Recommendations were also said to be associated with specialist terminology or were generic and therefore not relevant to specific personal situations. In addition, the probes showed that many participants were not familiar with any specific PA recommendations (e.g., from the World Health Organization [WHO]), but instead associated them with very generic statements, such as "you should move more/exercise is good/you should walk 10,000 steps a day." References to recommendations were therefore deleted from the items.

> *Because it depends on where the recommendation comes from. Does it really come from specialist staff? Is it easy to take? But if it really comes from people around you, then you do a bit of what you want. Or if someone is very ambitious and recommends going jogging for 5 hours a week, that just*

Table 4 Distribution of cognitive interview responses across QACS analysis categories, 2024

QACS category	Quantity	Percentage[a]
No. 7: Undefined/vague term	65	28.38
No. 20: Complex estimation, difficult mental calculation required	59	25.76
No. 18: High detail required or information unavailable	32	13.97
No. 4: Vague/unclear question	30	13.10
No. 6: Topic carried over from earlier question	22	9.61
No. 17: Shortage of memory cues	20	8.73
No. 21: Potentially sensitive or desirability bias	12	5.24
No. 10: Question too long	10	4.37
No. 8: Transition needed	9	3.93
Np. 11: Complex/awkward syntax	8	3.49
No. 19: Long recall of reference period	7	3.06
No. 5: Complex topic	3	1.31
No. 26: Missing response categories	3	1.31
No. 24: Responses use wrong or mismatching units	2	0.87
No. 9: Unclear respondent instruction	4	1.75

[a]Percentage: some items were coded with two QACS categories

doesn't make sense anymore. Then it's difficult. But it's difficult because the source of the recommendation is not entirely clear and you can't classify it. (Item 20, P18)

Another frequent QACS category was "Retrieval, No. 18: High detail required or information unavailable" (13.07%). The participants found the general term "diseases and complaints" used in the items from the healthcare dimension to be too imprecise. Both terms were subsequently defined more precisely in the introductory text as "complaints and diseases that affect you. This includes, for example, back pain, heart disease, joint problems, diabetes, and many others."

[…] the question, okay, [but] what is meant by illness, what is a physical complaint? (Item 1, P2)
I would define complaints. Because complaints are general, they can take any form. I would perhaps put an example behind it. (Item 3, P13)

The QACS category "Question content, No. 4: Vague/unclear question" (13.10%) was also considered important. In several items, for the dimension "applying," references to information ("based on information," e.g., "to decide based on information to use PA for the treatment of diseases and complaints") were perceived to be misaligned with individual lived experiences. Participants were more inclined to think about the active behavior (becoming physically active themselves) than the information they used for this purpose. In this context, the analytical process (guided by the joint discussions among the researchers involved) highlighted the need for a clear distinction between the new PA-HL measurement instrument and other related constructs, particularly health behavior. Consequently, and in line with a focus on information processing, it was decided to rename the dimension "applying," and all items associated with it, as "decision-making."

I don't use that at all, like magazines or online ads, to get active. Instead, I do it myself. I want to go out now, I want to go for a walk, I would like to ride a bike, or something like that. Honestly, in that moment, with that question, I wouldn't have thought about the doctor [or information]. I think it's also a bit about my mood on the day. (Item 19, P3)

Challenging level of abstraction (e.g., QACS No. 20)

The second most frequent QACS category was "Judgment and Evaluation, No. 20: Complex estimation, difficult mental calculation required" (25.76%). The quality of information (in all three dimensions) was perceived as important, and the significance of information quality and trustworthiness was repeatedly emphasized as being relevant to information processing. Subsequently, to assist participants in appraising potential information sources, examples for sources of information have been specified in the introductory text ("This information may originate from various sources, such as: conversations with healthcare professionals, (online) articles, videos, podcasts, informational brochures, or data from fitness trackers and smartphones. This also includes all information that you actively seek or passively receive.").

If someone writes something about physical exercise, then I think it's okay if they say, "Raise your right arm, that will help you to keep your shoulder joint in good

condition." Then I understand that. But if someone uses technical words, then of course I don't understand. Of course, it also depends on how these texts, this information, are written. (Item 14, P14)

Moreover, it became apparent that some participants were already assessing information quality or trustworthiness when responding to questions within the dimension "finding information." Thus, response behavior reflected not only the specific competency dimension "finding information" but overlapped with the dimension "appraisal" (e.g., whether the information was reliable/helpful). To avoid this overlap and confusion, corresponding instructions were subsequently provided in the introduction text, specifying that the questions should be answered with a focus on one dimension at a time. Additionally, relevant verbs within the items were underlined to serve as a visual reminder of the desired instruction.

I find it really difficult to find reliable information here. There's a lot of good advice, but I don't think all of it is actually effective. (Item 2, P6)

Missing content

None of the participants explicitly stated that specific content was missing, but during the course of the cognitive interviews, many participants mentioned the following topics: "ask professionals to explain content" (5/19 participants), "understand how to perform exercises during diseases" (6/19 participants), the issue of "coordinat[ing] time for PA" (6/19 participants), "find the right type of exercise" (5/19 participants), and "motivation" (7/19 participants). Based on this feedback, five new items concerning information processing competencies around these five topics were developed. Participants stated, for example, that:

[…] even if I don't interpret something like a knee injury as an illness, but just as a complaint, there are also differences as to how much is perhaps still good for you. And you can read all kinds of information about it, but I think you need expertise to decide for yourself […] (Item 8, P2)

Furthermore, the topic of the "regular implementation of PA," its feasibility in everyday life, and its challenging character was frequently mentioned (6/19 participants).

The continuity or discipline is then also to be done regularly. That is rather difficult. But could there be the addition [in the item] of perhaps implementing this regularly? (Item 9, P6)

To address this, some existing items were subsequently linguistically modified such that the qualifiers "regularly" or "sufficiently (frequently)" were integrated into certain items; for example, in item 11, "[…] to decide to regularly follow advice from professionals about physical activity when you have complaints or diseases in your daily life."

Other topics arose out of the data analysis from the cognitive interviews which were not considered for the adaptation process. The topic of "feeling the effects of PA on the body/body perception" is part of the construct of health consciousness (Hong 2009). The topic of "social influence" (or social interactions, e.g., a sense of team spirit or caregiving responsibilities for family members) plays a role in PA behavior, but it is considered as a determinant of HL, according to Sørensen et al. (2012, 2013). Other aspects (e.g., the difficulties in obtaining actual help due to long waiting times or short treatment duration) were considered not to be directly related to information processing competencies. In order to maintain clear conceptual delineation, therefore, these aspects were not considered for inclusion in the item pool.

Response categories

Nine of the 19 participants in the cognitive interviews stated that the response categories were easy to use and should not be changed (47.37%). For four participants, a "middle response category" was missing (21.05 %), and for a further three participants, the category "difficult" could not be sufficiently distinguished from "very difficult" (15.79%). One participant suggested alternative category wording (5.26%) ("very, partly, less, not at all"), and two participants would have preferred a six-point response scale (10.53%). This implies no clear trend regarding the response categories, although it can be said that nearly half of the participants were satisfied with the existing categories, and alternative suggestions only emerged through active inquiry. Since a "middle option" is deliberately avoided in order to inhibit neutral responses and enhance trend visibility, it was determined that the four-answer-option format would be maintained.

Synthesis of all results

The experts taking part in the eDelphi rounds were able to establish the basic structure and key topics for the PA-HL measurement instrument and significantly condense the item pool. The cognitive interviews provided valuable information on the understanding of items and content to be added. Feedback from the cognitive interviews and analysis according to the QACS, was used to linguistically revise all items. The concept of "information" posed challenges, particularly in the application dimension, where differentiating information from actual health behavior was not clearly understood by the participants. Subsequently, items relating to information processing competence were further refined. Based on the findings and further scientific approaches (DeGani et al.

2025; Robert Koch-Institut 2025; Wirtz and Soellner 2022), the definition of PA-HL was modified as follows:

> *PA-HL encompasses people's knowledge, motivation and competences to access, understand, appraise and decide upon PA-related information in different forms. As an information processing competence, PA-HL necessitates finding and understanding PA-related information, critically evaluating information on PA and making informed, health-related decisions to incorporate PA in everyday life with the goals of disease management, disease prevention and health promotion in order to maintain or improve quality of life over one's life course.*

The newly developed PA-HL measurement instrument includes 33 items and constitutes a subjective self-assessment. All items achieved sufficient content and face validity. Psychometric characteristics will be evaluated in a future pilot study (according to the study protocol) (Fitzner et al. 2025).

Discussion

This study describes a multistep process for developing and validating a PA-HL measurement instrument, incorporating an eDelphi process and cognitive interviews. The aim was to systematically gather feedback from experts and laypersons with regard to a PA-HL measurement instrument designed for adults in Germany. This methodological approach made it possible to develop and test an initial version of the instrument, which will undergo further psychometric testing in future studies. The eDelphi approach utilized a systematic consensus-building process to assess content validity, relevance parameters, and comprehensibility metrics of the novel PA-HL measurement instrument. Cognitive interviews were conducted to identify potential limitations and were analyzed utilizing the QACS. A significant methodological strength of this investigation is the integration of relevance assessment by experts with evaluation by lay participants in optimizing instrument utility.

Based on the study feedback, the PA-HL measurement instrument underwent substantial refinement in terms of both content and linguistic parameters. Findings from the eDelphi procedure indicated a conceptual differentiation between the disease prevention and health promotion domains, which could have been problematic for layperson comprehension. A consensus by experts stressed that measurement instruments designed for both scientific purposes and practical applications must be sufficiently simple in order to minimize participant loss during implementation. Recommendations were operationalized by simplifying linguistic structures and eliminating content redundancies. Over the course of three eDelphi iterations, the structural framework of the PA-HL measurement instrument underwent critical evaluation but was ultimately maintained in accordance with the framework proposed by Sørensen et al. (2012, 2013).

Analysis of the cognitive interviews revealed substantial comprehensibility of questionnaire items in terms of participants' immediate responses, based on the think-aloud technique. This technique has also proven beneficial in the development of other HL procedures such as that by Domanska et al. (2018) in examining the applicability of the German version of the European Health Literacy Survey Questionnaire (HLS-EU-Q47-GER) for adolescents, or by Singh et al. (2025) in assessing the comprehensibility of the 16-item version (HLS-EU-Q16) among long-term unemployed. Further modifications (including definitional clarifications, linguistic precision enhancements, and content supplementation) were undertaken based on these findings. Analysis of the cognitive interviews provides a thorough understanding of participants' experiences and difficulties with the PA-HL measurement instrument. A major difficulty for the participants was in the information processing in the competence area "applying," where participants did not necessarily reference the found/understood/appraised information. Referencing the conceptual framework proposed by Wirtz and Soellner (2022), the competence construct should be understood as a "guiding principle" rather than an explicit psychological construct, particularly with regard to operationalization and diagnostic assessments. Furthermore, the aim of this development process was to conceive a clearly delineated construct to avoid conceptual overlap and construct bias (Hair et al. 2019). As a result, the idea of refining the competence area "applying" underwent an iterative analysis process within the multidisciplinary research team. Consequently, the competence area "applying information" is specified as "decision-making."

This conceptual refinement is consistent with descriptions of the "applying" competency from other researchers. Wirtz and Soellner (2022) define it (Apply) as "communicating information and using it for health-enhancing decisions" (p. 11). DeGani et al. (2025) define it as "[making] informed decisions that positively impact their health and well-being" (p. 240). Furthermore, the German Robert Koch-Institut (2025) recently updated its definition of HL with a specific emphasis on decision-making: "The abilities to find, understand, evaluate, and apply health information for health-related decisions are referred to as 'health literacy'. This involves […] knowledge, motivation, and competences to form opinions and make decisions in everyday life […]." In their research on the information-seeking process, Heiberger et al. (2023) make a more explicit distinction between the components "apply" and "imaginated apply," the latter of which subsequently leads to decision-making. The explicit differentiation between actual and imagined application may enhance the specificity of HL models in general.

The conceptual refinement of decision-making involves an important methodological consideration ensuring a clear, content-related distinction between constructs relating to health behavior and those concerning motivational-volitional models like the health action process approach (HAPA) (Schwarzer et al. 2011). Health behavior encompasses any activity undertaken by an individual which prevents illness, helps detect it in an asymptomatic stage, averts damage, or extends life expectancy (Lippke and Renneberg 2006; Schwarzer et al. 2011). This distinction is especially relevant for procedures of instrument development, including steps undertaken to assess construct validity, such as exploratory and confirmatory factor analyses. Conceptual overlap can lead to, among other things, cross-loadings, reduced discriminant validity, or poorly fitted models (Brown 2015; Kline 2023). Wirtz and Soellner (2022) argue that inadequate delineation of HL has left essential questions regarding content validity, construct validity, and the utility of measurement outcomes unresolved. DeGani et al. (2025) advocate for a precise and standardized conceptualization of HL. They establish a clear distinction between HL itself and "HL promotion," and treat them as separate domains. Only the latter term encompasses the active engagement with health-related challenges and barriers.

According to Gugglberger (2019), HL and health promotion have a multifaceted relationship, whereby HL can serve as the foundation, driver, or partner of health promotion. HL can be conceptualized as either a moderating or mediating variable in the complex interrelationships between various health determinants (Pelikan et al. 2018; Schillinger 2020). Accordingly, PA-HL could potentially function as a determinant of health behavior. Nevertheless, it is important to emphasize that in our conceptual understanding of PA-HL, the dimension of application does not refer to behavior. Health behavior, in this context, is a distinct construct with distinct measurement instruments, and should not be conflated with HL. In summary, the definition of PA-HL has been revised in the measurement instrument to explicitly clarify that the "application of information" is not intended to be understood in the sense of behavior.

Regarding the response categories and associated behavior, it should be emphasized that the PA-HL measurement instrument is designed as a self-assessment tool, which inherently cannot provide objective insights. The authors acknowledge that the concept of HL, and consequently that of PA-HL, necessitates that critical reflection and interpretation be undertaken with caution and contextual considerations. Accordingly, PA-HL is understood as a "self-reported performance" of the participants, and any evaluation should occur within a broader discourse regarding the advantages and limitations of self-assessment methodologies. An advantage of self-assessment tools is that they provide a feasible approach for dealing with complex constructs. This feasibility, however, means that the competence of the user in terms of their self-assessment skills must also be considered (Wirtz and Soellner 2022), and highlights a potential disadvantage. This required user competency is due to factors such as the influence of their perception of health-related behavior, social desirability, and other significant factors (Wirtz and Soellner 2022). Despite methodological constraints inherent to self-assessment protocols, this PA-HL construct provides added value to the scientific discourse in HL research. On the one hand, it adopts the conceptualization of Sørensen et al. (2012, 2013), contributing to an understanding of HL as an "information-processing competence," and distinguishing it from other constructs. On the other hand, it critically reflects on its underlying conceptualization and offers responses to its weaknesses.

Strengths and limitations

Despite the numerous methodological strengths (e.g., a combined approach involving expert consultation and testing on a target population), some limitations of the methods used here need to be acknowledged. Notwithstanding the anonymous nature of the first and third online rounds in the eDelphi process, participants may have been biased in their responses by social desirability, potentially perceiving that elevating ratings was professionally expected (Niederberger and Renn 2019). Moreover, all eDelphi data were derived from availability sampling and exclusively reflect subjective expert opinions and specialized knowledge (consistent with the inherent limitations of Delphi methodologies). This was mitigated by the inclusion of an interdisciplinary team during the eDelphi process, incorporating diverse professional perspectives and an appropriate number of participants in all data collection rounds.

In the cognitive interviews, participants often had to remember past thoughts or decision-making processes (Conrad and Blair 2009). This retrospectivity can lead to distortions, where interviewees may inaccurately reconstruct or rationalize their thoughts. The verbalization of introspective thought processes represented an atypical cognitive task for many participants, potentially resulting in the sharing of either incomplete or overly elaborate thought processes. It is also recognized that probing is susceptible to reactivity bias, in which respondents modify their thought processes merely because they are asked to articulate them (Conrad and Blair 2009). For example, an interviewer probing the interpretation of specific terminology may direct respondents to consider various potential (mis)interpretations of items, even when their initial perception indicated unambiguous comprehension (Lenzner et al. 2024). Although this study comprised 19 participants with a targeted recruitment of demographic diversity in terms of age, sex, migration background, and educational attainment, the final sample had a slightly higher percentage of more highly educated persons. Additionally, only four participants reported having a migration background, which

also decreases the heterogeneity of the sample. However, such methodological constraints are well documented, and the think-aloud technique remains a recommended protocol for cognitive interviews. In addressing these concerns, we considered the substantial advantages of cognitive interviewing, including the use of think-aloud techniques and probing (Brancato et al. 2006; Knafl et al. 2007; Pohontsch and Meyer 2015). By identifying misinterpretations at an early stage, questionnaire items were revised throughout the instrument development process. This enhanced the content alignment between questions and the construct to be measured. Furthermore, the number of participants recruited for the cognitive interviews was at the upper end of recommended sample sizes (Lenzner et al. 2024; Pohontsch and Meyer 2015; Willis 2005), and the entire process was documented transparently (see supplementary material).

Conclusions

Individuals require specific competencies for processing PA-relevant information, accessing services, and addressing challenges in order to make informed and sustainable decisions regarding health and well-being, not only for themselves but also for others. Consequently, the competencies necessary for dealing with PA-related information in a meaningful and appropriate manner have acquired significant relevance in contemporary times. The development of this PA-HL measurement instrument represents an attempt to contribute to understanding the facilitators and barriers in terms of future population health. The implementation of an eDelphi technique and cognitive interviews to identify and establish a consensus on which PA-HL items were to be included and/or modified afforded an evidence-based approach to the development of the measurement instrument. The integration of this PA-HL measurement instrument within future research investigations, healthcare applications, or municipal and regional surveys represents a novel approach to public health that has the potential to help investigators better understand the barriers in terms of the processing competencies of adults regarding PA information. PA-HL measurement instrument outcomes could aid in the optimization of future information dissemination strategies, potentially influencing health behavior patterns and contributing to NCD prevention initiatives. A psychometric evaluation of the PA-HL measurement instrument is planned in order to establish validity and reliability parameters for future implementation.

Abbreviations eDelphi: Online Delphi; HL: Health literacy; I-CVI: Item content validity index; NCD: Noncommunicable disease; PA: Physical activity; PA-HL: Physical activity-related health literacy; S-CVI/Ave: Average scale content validity index; QACS: Questionnaire appraisal coding scheme

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10389-026-02683-6.

Acknowledgements We would like to acknowledge all the participants who participated in the eDelphi and cognitive interviews.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability The datasets used during the study will be available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate The study was designed following the principles of the Declaration of Helsinki. Ethical approval was granted by the independent ethics committee of the University of Applied Sciences Fulda on December 21, 2023, and June 21, 2024 (reference EK231204), which included participant information material, measurement instruments, and an informed consent form.

Consent for publication All participants provided consent for the use of anonymized data for publication; no personal data will be published.

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

Clinical trial number Not applicable. No clinical trial.

References

Abioye AI, Hajifathalian K, Danaei G (2013) Do mass media campaigns improve physical activity? A systematic review and meta-analysis. Arch Public Health 71:1–10. https://doi.org/10.1186/0778-7367-71-20

Beatty PC, Willis GB (2007) Research synthesis: the practice of cognitive interviewing. Public Opin Q 71(2):287–311. https://doi.org/10.1093/poq/nfm006

Benda F (2019) Strukturelle Prädiktoren der Adhärenz für körperliche Aktivität im betrieblichen Setting – eine systematische Übersichtsarbeit. B&G Bewegungstherapie und Gesundheitssport 35(03):127–133. https://doi.org/10.1055/a-0890-7269

Bethmann A, Buschle C & Reiter H (2019). Kognitiv oder qualitativ? In N. Menold (Ed.), Schriftenreihe der asi - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute Ser. Qualitätssicherung Sozialwissenschaftlicher Erhebungsinstrumente (pp. 159–193). Vieweg. https://doi.org/10.1007/978-3-658-24517-7_6

Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL (2018) Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 6:149. https://doi.org/10.3389/fpubh.2018.00149

Bonnar-Kidd KK, Black DR, Mattson M, Coster D (2009) Online physical activity information: will typical users find quality information? Health Commun 24(2):165–175. https://doi.org/10.1080/10410230802676763

Brancato G, Macchia S, Murgia M, Signore M, Simeoni G, Blanke K & Hoffmeyer-Zlotnik J (2006). Handbook of recommended practices for questionnaire development and testing in the European statistical system. European Statistical System.

Brown TA (2015) Confirmatory factor analysis for applied research. The Guilford Press

Buecker S, Simacek T, Ingwersen B, Terwiel S, Simonsmeier BA (2021) Physical activity and subjective well-being in healthy individuals: a meta-analytic review. Health Psychol Rev 15(4):574–592. https://doi.org/10.1080/17437199.2020.1760728

Buja A, Rabensteiner A, Sperotto M, Grotto G, Bertoncello C, Cocchio S, Baldovin T, Contu P, Lorini C, Baldo V (2020) Health literacy and physical activity: a systematic review. J Phys Act Health 17(12):1259–1274. https://doi.org/10.1123/jpah.2020-0161

Bundesvereinigung Prävention und Gesundheitsförderung e.V. (Ed.) (2025). Herausforderungen und Chancen zur Weiterentwicklung von Prävention und Gesundheitsförderung in der 21. Legislaturperiode: Policy Paper der Bundesvereinigung Prävention und Gesundheitsförderung e.V. (BVPG). https://bvpraevention.de/newbv/images/Publikationen/BVPG_Policy_Paper_2025.pdf

Carl J, Domanska OM, Buchmann M, Jordan S, Pfeifer K, Sudeck G (2024) Toward a representative assessment of competencies for a healthy, physically active lifestyle in the German population: development and validation of a short version of the physical activity-related health competence questionnaire (PAHCO_12). Germ J Exerc Sport Res. https://doi.org/10.1007/s12662-024-00943-9

Chinn D, McCarthy C (2013) All aspects of health literacy scale (AAHLS): developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. Patient Educ Couns 90(2):247–253. https://doi.org/10.1016/j.pec.2012.10.019

Conrad FG, Blair J (2009) Sources of error in cognitive interviews. Public Opin Q 73(1):32–55. https://doi.org/10.1093/poq/nfp013

DeGani SM, Alder E, Beese AS (2025) Das Konzept Gesundheitskompetenz im Wandel. Bundesgesundheitsblatt Gesundheitsforsch Gesundheitsschutz 68(3):240–246. https://doi.org/10.1007/s00103-024-04004-2

DeMaio TJ & Landreth A (2004). Do different cognitive interview techniques produce different results? Methods for Testing and Evaluating Survey Questionnaires, 89–108. https://doi.org/10.1002/0471654728.ch5

Domanska OM, Firnges C, Bollweg TM, Sørensen K, Holmberg C, Jordan S (2018) Do adolescents understand the items of the European Health Literacy Survey Questionnaire (HLS-EU-Q47)–German version? Findings from cognitive interviews of the project "Measurement of Health Literacy Among Adolescents"(MOHLAA) in Germany. Arch Public Health 76:1–14. https://doi.org/10.1186/s13690-018-0276-2

Ericsson KA, Simon HA (1980) Verbal reports as data. Psychol Rev 87:215–250. https://doi.org/10.1037/0033-295X.87.3.215

Finbråten HS, Nowak P, Griebler R, Bíró É, Vrdelja M, Charafeddine R, Griese L, Bøggild H, Schaeffer D, Link T (2022) The HLS19-COM-P, a new instrument for measuring communicative health literacy in interaction with physicians: development and validation in nine European countries. Int J Environ Res Public Health 19(18):11592. https://doi.org/10.3390/ijerph191811592

Fitzner E, Hering T, Dadaczynski K (2025) Development and validation of a measurement instrument for physical activity-related health literacy (PA-HL): a study protocol. Arch Public Health 83(1):165. https://doi.org/10.1186/s13690-025-01661-w

Froböse I & Wallmann-Sperlich B (2025). Der DKV-Report 2025 | Wie gesund lebt Deutschland? https://www.dkv.com/downloads/DKV-Report-2025-Bericht.pdf

Garcia L, Pearce M, Abbas A, Mok A, Strain T, Ali S, Crippa A, Dempsey PC, Golubic R, Kelly P, Laird Y, McNamara E, Moore S, de Sa TH, Smith AD, Wijndaele K, Woodcock J, Brage S (2023) Non-occupational physical activity and risk of cardiovascular disease, cancer and mortality outcomes: a dose–response meta-analysis of large prospective studies. Br J Sports Med 57(15):979–989. https://doi.org/10.1136/bjsports-2022-105669

Goodyear VA, Boardley I, Chiou SY, Fenton S, Makopoulou K, Stathi A, Wallis GA, van Veldhuijzen Zanten JJCS, Thompson JL (2021) Social media use informing behaviours related to physical activity, diet and quality of life during COVID-19: a mixed methods study. BMC Public Health 21:1–14. https://doi.org/10.1186/s12889-021-11398-0

Goto E, Ishikawa H, Nakayama K, Kiuchi T (2018) Comprehensive health literacy and health-related behaviors within a general Japanese population: differences by health domains. Asia Pac J Public Health 30(8):717–726. https://doi.org/10.1177/1010539518806806

Griebler R, Link T, Schütze D, Straßmayr C (2025) Gesundheitskompetenz messen: Methoden und Instrumente zur Erfassung der allgemeinen Gesundheitskompetenz bei Erwachsenen. Bundesgesundheitsblatt Gesundheitsforsch Gesundheitsschutz 68(3):247–254. https://doi.org/10.1007/s00103-025-04010-y

Griese L, Schaeffer D (2022) Die Navigation im Gesundheitssystem erleichtern – erste Daten zur navigationalen Gesundheitskompetenz. Public Health Forum 30(2):135–137. https://doi.org/10.1515/pubhef-2022-0020

Gugglberger L (2019) The multifaceted relationship between health promotion and health literacy. Health Promot Int 34(5):887–891. https://doi.org/10.1093/heapro/daz093

Häder M (2009) Delphi-Befragungen: Ein Arbeitsbuch. Springer. https://doi.org/10.1007/978-3-658-01928-0

Hair F, Black W, Babin B, Anderson R (2019) Multivariate Data Analysis, 8th ed. Cengage Learning

Heiberger A, Dresch C, Schulz AA, Wirtz MA (2023) Health literate internet-based information-seeking processes: theory-based development of a conceptual model. J Med Internet Res 25:e39024. https://doi.org/10.2196/39024

Hong H (2009). Scale development for measuring health consciousness: Re-conceptualization. That Matters to the Practice, 212. https://www.researchgate.net/profile/Bruce-Berger-7/publication/254392294_Role_Modeling_in_Public_Relations_The_Influence_of_Role_Models_and_Mentors_On_Leadership_Beliefs_and_Qualities/links/546deced0cf2a7492c56d399/Role-Modeling-in-Public-Relations-The-Influence-of-Role-Models-and-Mentors-On-Leadership-Beliefs-and-Qualities.pdf#page=212

Idowu OA, Omoijahe B, Fawole HO, Adeagbo I, Akinola BI (2024) Physical activity information-seeking behaviour and barriers in a sample of university undergraduate emerging adults: a cross-sectional survey. Bull Fac Phys Ther 29(1):20. https://doi.org/10.1186/s43161-024-00187-6

Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG (2017) Guidance on conducting and reporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological

systematic review. Palliat Med 31(8):684–706. https://doi.org/10.1177/0269216317690685
Kaňková J, Binder A, Matthes J (2024) Helpful or harmful? Navigating the impact of social media influencers' health advice: insights from health expert content creators. BMC Public Health 24(1):3511. https://doi.org/10.1186/s12889-024-21095-3
Kline RB (2023) Principles and practice of structural equation modeling. The Guilford Press
Knafl K, Deatrick J, Gallo A, Holcombe G, Bakitas M, Dixon J, Grey M (2007) The analysis and interpretation of cognitive interviews for instrument development. Res Nurs Health 30(2):224–234. https://doi.org/10.1002/nur.20195
Köck-Hódi S, Mayer H (2013) Die Delphi-Methode. Procare 18(5):16–20. https://doi.org/10.1007/s00735-013-0094-2
Lee J, Turner K, Xie Z, Kadhim B, Hong YR (2022) Association between health information-seeking behavior on YouTube and physical activity among US adults: results from health information trends survey 2020. AJPM Focus 1(2):1–8. https://doi.org/10.1016/j.focus.2022.100035
Lenzner T, Hadler P & Neuert CE (2024). Cognitive Pretesting. GESIS., Article 3. Advance online publication. https://doi.org/10.15465/gesis-sg_en_049
Lippke S & Renneberg B (2006). Theorien und Modelle des Gesundheitsverhaltens. In B. Renneberg & P. Hammelstein (Eds.), Gesundheitspsychologie (pp. 35–60). Springer-Verlag. https://doi.org/10.1007/978-3-540-47632-0_5
Lynn MR (1986) Determination and quantification of content validity. Nurs Res 35(6):382–386
Matterne U, Apfelbacher C, Hasenpusch C (2022) Gesundheitskompetenz in der Allgemeinbevölkerung im Kontext der COVID-19 Pandemie. Public Health Forum. https://doi.org/10.1515/pubhef-2022-0032
Mayring P & Fenzl T (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), Handbuch Methoden der empirischen Sozialforschung (pp. 543–556). Springer VS. https://doi.org/10.1007/978-3-658-21308-4_42
Niederberger M, Renn O (2019) Delphi-Verfahren in den Sozial- und Gesundheitswissenschaften: Konzept. Springer, Varianten und Anwendungsbeispiele. https://doi.org/10.1007/978-3-658-21657-3
Pelikan JM, Ganahl K, Roethlin F (2018) Health literacy as a determinant, mediator and/or moderator of health: empirical models using the European Health Literacy Survey dataset. Glob Health Promot. https://doi.org/10.1177/1757975918788300
Pohontsch N, Meyer T (2015) Das kognitive Interview - Ein Instrument zur Entwicklung und Validierung von Erhebungsinstrumenten. Rehabil 54(1):53–59. https://doi.org/10.1055/s-0034-1394443
Polit DF, Beck CT, Owen SV (2007) Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 30(4):459–467. https://doi.org/10.1002/nur.20199
Prüfer P & Rexroth M (2005). Kognitive Interviews [https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201470]. ZUMA How-to-Reihe. http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/how_to15pp_mr.pdf 201470
Rhodes RE, Quinlan A (2015) Predictors of physical activity change among adults using observational designs. Sports Med 45(3):423–441. https://doi.org/10.1007/s40279-014-0275-6
Robert Koch-Institut (2025). Gesundheitskompetenz / Health Literacy. https://www.rki.de/DE/Themen/Gesundheit-und-Gesellschaft/Gesundheitliche-Einflussfaktoren-A-Z/G/Gesundheitskompetenz/gesundheitskompetenz-node.html
Schaber K, Leitzmann M, Pukrop T, Quentin W, Stein MJ, Jochem C (2025) Burden of cancer attributable to sedentary behaviour in Germany: an epidemiological analysis of survey data. BMJ Open. https://doi.org/10.1136/bmjopen-2024-094304
Schaeffer D, Gille S, Berens EM, Griese L, Klinger J, Vogt D, Hurrelmann K (2021) Digitale Gesundheitskompetenz der Bevölkerung in Deutschland Ergebnisse des HLS-GER 2 [Digital Health Literacy of the Population in Germany: Results of the HLS-GER 2]. Advance online publication, Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). https://doi.org/10.1055/a-1670-7636
Schillinger D (2020) The intersections between social determinants of health, health literacy, and health disparities. Stud Health Technol Inform 269:22–41. https://doi.org/10.3233/SHTI200020
Schladitz L (2019). Gesundheitskompetenz der deutschsprachigen Patienten im zahnmedizinischen Setting [, Westsächsische Hochschule Zwickau]. EndNote Tagged Import Format.
Schwarzer R, Lippke S, Luszczynska A (2011) Mechanisms of health behavior change in persons with chronic illness or disability: the health action process approach (HAPA). Rehabil Psychol 56(3):161. https://doi.org/10.1037/a0024509
Singh H, Härtling V, Samkange-Zeeb F, Kolschen J, Brand T, Zeeb H, Schüz B (2025) Comprehensibility of the HLS-EU-Q tool for assessing health literacy of long-term unemployed individuals: A qualitative study using the think-aloud method. Advance online publication, Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)). https://doi.org/10.1055/a-2570-0424
Sørensen K, van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H (2012) Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 12:80. https://doi.org/10.1186/1471-2458-12-80
Sørensen K, van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, Kondilis B, Stoffels V, Osborne RH, Brand H (2013) Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 13(1):1–10
Strain T, Flaxman S, Guthold R, Semenova E, Cowan M, Riley LM, Bull FC, Stevens GA (2024) National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5· 7 million participants. Lancet Glob Health 12(8):e1232–e1243. https://doi.org/10.1016/S2214-109X(24)00150-5
Sudeck G, Rosenstiel S, Carl J & Pfeifer K (2022). Bewegungsbezogene Gesundheitskompetenz–Konzeption und Anwendung in Gesundheitsförderung, Prävention und Rehabilitation. In Gesundheitskompetenz (pp. 1–12). Springer. https://doi.org/10.1007/978-3-662-62800-3_135-1
Tavousi M, Haeri-Mehrizi A, Rakhshani F, Rafiefar S, Soleymanian A, Sarbandi F, Ardestani M, Ghanbari S, Montazeri A (2020) Development and validation of a short and easy-to-use instrument for measuring health literacy: the health literacy instrument for adults (HELIA). BMC Public Health 20(1):656. https://doi.org/10.1186/s12889-020-08787-2
Thomas JD, Cardinal BJ (2020) How credible is online physical activity advice? The accuracy of free adult educational materials. Transl J Am Coll Sports Med 5(9):82–91. https://doi.org/10.1249/TJX.0000000000000122
Tong A, Sainsbury P, Craig J (2007) Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 19(6):349–357
Tourangeau R, Rips LJ, Rasinski K (2000) The psychology of survey response. Cambridge University Press
Waltz CF, Strickland OL & Lenz ER (2010). Measurement in nursing and health research (Fourth Edition). Springer publishing company.

Warburton DE, Bredin SS (2017) Health benefits of physical activity: a systematic review of current systematic reviews. Curr Opin Cardiol 32(5):541–556. https://doi.org/10.1097/HCO.0000000000000437

Willis GB (2005) Cognitive interviewing: A tool for Improving Questionnaire Design. sage publications

Wirtz M, Soellner R (2022) Gesundheitskompetenz – Konstruktverständnis und Anforderungen an valide Assessments aus Perspektive der psychologischen Diagnostik. Diagnostica 68:163–171. https://doi.org/10.1026/0012-1924/a000299

World Health Organization (2024). Physical Activity. https://www.who.int/news-room/fact-sheets/detail/physical-activity

Xie Y, An W, Zheng C, Zhao D, Wang H (2022) Multidimensional health literacy profiles and health-related behaviors in the elderly: a cross-sectional survey. Int J Nurs Sci 9(2):203–210. https://doi.org/10.1016/j.ijnss.2022.02.007