



Use of Item Response Models in Assessing the Health Literacy Facet Understanding Health Information for Early Childhood Allergy Prevention and Prevention of COVID-19 Infections by Pregnant Women and Mothers of Infants

Anja Alexandra Schulz^{ID}, Carolin Dresch^{ID}, Andrea Heiberger^{ID}, and Markus Antonius Wirtz^{ID}

Research Methods in Health Sciences, Faculty for Mathematics, Natural Sciences and Technology, University of Education Freiburg, Germany

Abstract: Appropriate parental health literacy (HL) is essential to preventively maintain and promote child health. Understanding health information is assumed to be fundamental in HL models. We developed $N = 67$ items (multiple-choice format) based on information materials on early childhood allergy prevention (ECAP) and prevention of COVID-19 infections to assess the parental HL facet Understand. $N = 343$ pregnant women and mothers of infants completed the items in an online assessment. Using exploratory factor analysis for ordinal data (RML estimation) and item response models (1-pl and 2-pl model), we proved the psychometric homogeneity of the item pool. 57 items assess the latent dimension Understand according to the assumptions of the 1-pl model (weighted MNSQ < 1.2 ; separation reliability = .855). Person parameters of the latent trait Understand correlate specifically with subjective socioeconomic status ($r = .27$), school graduation ($r = .46$), allergy status ($r = .11$), and already infected with COVID-19 ($r = .12$). The calibrated item pool provides a psychometrically sound, construct-valid assessment of the HL facet Understand Health Information in the areas of ECAP and prevention of COVID-19 infections.

Keywords: parental health literacy, understanding health information, prevention, literacy structure modeling, item response models

Anwendung von Item-Response-Modellen bei der Erfassung der Gesundheitskompetenzfacette Verstehen von Gesundheitsinformationen zur frühkindlichen Allergieprävention und der Prävention von COVID-19-Infektionen bei Schwangeren und Müttern von Kleinkindern

Zusammenfassung: Hinreichende elterliche Gesundheitskompetenz (GK) ist wesentlich, um die Gesundheit des Kindes erhalten und präventiv fördern zu können. Dem angemessenen Verstehen von Gesundheitsinformationen kommt in GK-Modellen eine grundlegende Funktion zu. Für die beiden Anwendungsgebiete frühkindliche Allergieprävention (FKAP) und Prävention von COVID-19 Infektionen wurden $N = 67$ Testitems auf Basis von Informationsmaterialien entwickelt. $N = 343$ Schwangere und Mütter von Kleinkindern bearbeiteten die Items in einer Online-Testung. 57 Testitems erfassen die latente Dimension Verstehen gemäß des 1-pl Item Response Modells (Weighted MNSQ $< 1,2$; Separations-Reliabilität = .874). Die Personenparameter der latent trait Verstehen von Gesundheitsinformationen korrelieren u. a. mit dem subjektiven sozioökonomischen Status ($r = .27$), dem Schulabschluss ($r = .46$), dem Allergiestatus ($r = .11$) und bereits erfolgter COVID-19-Infektion ($r = .12$) der Frauen. Die kalibrierte Itemgruppe erlaubt eine psychometrisch fundierte, konstruktvalid Erfassung der GK-Facette Verstehen von Gesundheitsinformation in den Anwendungsgebieten FKAP und Prävention von COVID-19-Infektionen.

Schlüsselwörter: Elterliche Gesundheitskompetenz, Verstehen von Gesundheitsinformationen, Prävention, Kompetenzstrukturmodellierung, Item-Response-Modelle

Parental health literacy (HL) is essential in the protection of infant health and to preventively promote long-term

health development in children (DeWalt & Hink, 2009). HL is an individual resource that can be influenced and

promoted through learning and experience. While engagement with the concept of HL is concerned with the individual's own health (Nutbeam, 2000; Soellner et al., 2010; Sørensen et al., 2012), parental HL focuses on the parents' healthcare for the developing and growing child. Parents have to make health-literate decisions and implement appropriate behaviors, with the goal of maintaining their child's health and promoting healthy development. Parental HL has been studied primarily for chronically ill children (curative or tertiary preventive behaviors; DeWalt & Hink, 2009). Low parental HL was found to be associated with low health-related knowledge, detrimental or harmful behaviors to the child, and a lack of understanding and implementation of health-promoting behaviors for the child (Borges et al., 2017). Existing research applies assessment instruments (especially self-assessments; Sørensen et al. 2013), which fail to meet the psychometric criteria for sound measurement of competencies (objective assessments according to item response theory; IRT, Rupp et al., 2010). Hence, the following emphasizes that (a) a valid definition and (b) an appropriate structural model of the HL-construct as well as (c) a psychometrically sound assessment of its facets are fundamental for research on the HL construct (Dresch et al., 2021).

Health Literacy – Definition and Models

According to Nutbeam's (2000) definition, health literate people have cognitive, social, and motivational skills and abilities that enable them to identify, process, and understand health information and services to make appropriate decisions and act on that information. Functional HL refers to literacy skills and abilities required for validly reading and understanding health information (including numerical skills; Numeracy; Nutbeam, 2000; Sørensen et al., 2012). Communicative or Interactive HL is required for goal-oriented behavior in the healthcare system, and Critical HL enables people to appropriately evaluate health information (Nutbeam, 2000). In addition to literal skills, health-related knowledge is defined as a further fundamental component of HL in Soellner and colleagues' model (2010). Particular importance is attached to motivational and action-oriented facets, which play a major role in determining the health-literate actions of individuals. Sørensen et al. (2012) assume that the successive steps of Access, Understand, Appraise, and Apply must be regarded as central facets of HL. All HL models share in common that (1) HL must be understood as a multidimensional construct and (2) literate abilities and skills – in the sense of Functional HL – represent fundamental model facets as well as determinants of health-literate behavior.

The models differ primarily regarding the consideration of further psychological and systemic constructs as well as assumptions about decision-making and behavioral processes (Mackert et al., 2015; Soellner & Rudinger, 2018; Sørensen et al., 2012).

Developing an Assessment for the HL Facet Understand Health Information

In this article, methods of competence structure and level modeling are used for the investigation and measurement of the HL facet Understand Health Information (Klieme et al., 2008). Although this is only one facet of the HL construct, it is defined as a distinct subconstruct and is included in all models as a fundamental requirement for health-literate information processing. The facet Understand is most likely to be considered in the sense of Functional HL (Nutbeam, 2000). Furthermore, a valid understanding of information is a necessary prerequisite for health-literate information appraisal and decision-making (Paasche-Orlow & Wolf, 2007). Since an achievement test is to be developed for the living situation of new parents, a clear distinction has to be made here – independent of personal and social moderator effects – between correct and incorrect results of cognitive information processing. Because of the clarity of the response scoring and the definition of distinct elementary processing requirements using single items, this opens up a solid basis for transferring the standards of competence modeling (item response models) to the field of health competence modeling (Klieme et al., 2008; Pant, 2013; Schmuckle & Egloff, 2011).

Competencies or literacy traits are generally defined as personal behavioral dispositions that can be actualized or utilized purposefully in information processing as well as in planning and controlling behavior. Competencies determine the successful solution or accomplishment of performance requirements or tasks (Weinert, 2001). By definition, competencies are application-oriented and domain-specific (Weinert, 2001). Competence levels are determined in particular by acquired knowledge and learning experiences. Accordingly, a valid survey on HL has to be situation- and context-related to ensure that parents' daily life conditions and the ensuing demands are considered. To meet these requirements and to be able to investigate domain specificity of Understand Health Information, the present study strived to develop parental HL items based on information materials targeted to parents for two content areas: early childhood allergy prevention (ECAP; Dresch et al., 2021) and COVID-19 infection prevention (COVID-19-IP; Cluver et al., 2020).

Health information materials can be considered a medium intended to induce certain cognitive changes in the recipients (e.g., enhanced problem awareness, knowledge or problem-solving skills) and, consequently, to improve their health-related decision-making and actions. Health-literate people possess the cognitive abilities to extract the right information from health information or draw the right conclusions from the materials for themselves. Faulty understanding occurs when information is derived from materials that contradict the intentions of the media designer. However, erroneous understanding is not necessarily caused by erroneous processing; the media content itself can be systematically error-prone or induce misconceptions (Andrulis & Brach, 2007).

Accordingly, the validity of the content and the didactic design must first be ensured (Wilson, 2005). The following aspects must be considered for an objective performance measure of Understand Health Information (Dresch et al., 2021):

1. Does the information material belong in the defined content area (ECAP, COVID-19-IP)? Is it representative of the content area?
2. Does the material aim to change cognitive representations in people who do not yet have the appropriate knowledge (especially misconceptions)?
3. Is the information to be conveyed clearly defined?
4. Is the information to be conveyed evident or suitable to validly increasing health-related knowledge?
5. Is the presentation format appropriate to conveying the information adequately? Can erroneous processing result from the design of the representation?
6. Is the information material suitable for the target group? What prior knowledge is assumed? Are the information needs and the life situation of the target group taken into account?
7. Can response options be identified that clearly indicate whether the content to be conveyed was validly processed / understood by the respondents (target responses) or was incorrectly processed / understood (distractors)?

Psychometric Modeling and Measurement Standards for Competence / Literacy Assessment

Particularly in the field of educational research, methods of competence structure and competence level modeling based on the assumptions of item response theory (IRT; Rupp et al., 2010; van der Linden, 2021) represent the diagnostic gold standard for analyzing competency domains and capturing competence levels (Wilson, 2005). However, methods for assessing HL do not yet use these advantageous psychometric standards (Haun et al., 2014).

Subjective assessments (e.g., European Health Literacy Survey Questionnaire [HLS-EU-Q], Sørensen et al. 2013) capture the self-assessed ability to identify and process health information and to decide and act on this basis. Measures obtained by self-assessments systematically deviate empirically and conceptually from objective indicators of competence because of individual information processing (e.g., individual reference systems, self-serving bias; Schmuckle & Egloff, 2011). Existing objective achievement tests (e.g., Test of Functional Health Literacy in Adults [TOFHLA]; Parker et al., 1995) are generally limited to basal literacy skills (word recognition, reading comprehension, numeracy). Additionally, the psychometric foundations of existing objective test procedures are limited to basal validations according to the classical test theory and limited psychometrically reasonable evaluation rules (Eid & Schmidt, 2014).

The present paper examines whether the basic HL facet Understand Health Information can be modeled as a latent trait dimension by items based on informational materials targeting pregnant women and mothers of infants. IRT models (van der Linden, 2001; Wilson, 2005) assume that the solution probability $p(+)$ results in a logistic transformation of the difference in ability θ_v of a woman v , the difficulty σ_i of the task i to be solved, and its discrimination α_i :

$$p(+|\theta_v, \sigma_i, \alpha_i) = \frac{e^{\alpha_i \cdot (\theta_v - \sigma_i)}}{1 + e^{\alpha_i \cdot (\theta_v - \sigma_i)}}$$

If the discrimination strength is identical for all items ($\alpha_i = 1$; $\text{VAR}(\alpha_i) = 0$), this model is called the 1-pl or the Rasch model. In the case of the 2-pl or Birnbaum model, the discriminatory strengths of the items are allowed to vary and are accordingly estimated item-specifically ($\text{VAR}(\alpha_i) \neq 0$). Item groups satisfying the IRT assumptions form an ideal basis for analyzing the dimensional structure of literacy domains, since the assumption that the underlying latent dimensions – in this case, Understand Health Information – correspond unambiguously with the empirical data information is tested empirically. If it is assumed that not only one but multiple competence dimensions or facets determine the solution of tasks, multidimensional IRT models (mIRT; Rupp et al., 2010) can be used.

The central aim of this study is to test the psychometric and dimensional structure of an item pool for Understand Health Information. Using IRT methods, we investigate the following questions based on test items developed for the application areas of ECAP and COVID-19-IP by pregnant women and mothers of infants.

1. Is there a common latent HL dimension Understand Health Information underlying the processing of the test items?
2. Does the model fit improve by distinguishing the prevention contents (1) ECAP, (2) COVID-19-IP, and (3) General Health Prevention (GHP)?
3. Which items represent the underlying HL dimension(s) according to the assumptions of IRT?
4. Are the identified HL levels of Understand Health Information associated with sociodemographic and allergy- and COVID-19-related characteristics of pregnant women and mothers of infants?

Methods

The study was conducted as part of the DFG Research Group “Health Literacy in Early Childhood Allergy Prevention: Parental Competencies and Public Health Context in a Shifting Evidence Landscape” [FOR 2959; G.Z.: AP 235/3-1] within the subproject “Structural Modelling and Assessment of Health Literacy in Allergy Prevention of New Parents” [G.Z.: WI-3210/7-1]. The German Psychological Society gave a positive ethical vote (registration no. MAW 112018), and the study was preregistered at the Leibniz Institute of Psychology (ZPID) (Wirtz et al., 2021).

Item Construction

The definition of item contents was based on a comprehensive literature and internet search on health information in ECAP and COVID-19-IP. The focus was on information directly targeted to parents in the area of ECAP and adults in general in COVID-19-IP (e.g., www.allergiecheck.de, www.quarks.de, www.gesundheitsinformation.de, www.rki.de).

Additionally, 13 parents and 5 students conducted an internet-based search on health-related information on ECAP (average duration approximately 120 min; range: 41–252 min). The screen contents and the continuous accompanying verbalization of parental thoughts (think-aloud method; Pohontsch & Meyer, 2015) were recorded. In a subsequent guided interview (average duration: 43 min, range: 35–60 min), the parents reflected in particular on their prior knowledge and information needs, their search strategies, and their assessment of the quality and usefulness of the identified information materials. Based on this, an item pool with 67 test items was defined and item-specific response options for the multiple-choice answer format (correct answers and distractors corresponding to typical comprehension errors) were used.

Each item consists of an item stem (graph, table, text) conveying information and answer alternatives, one or more of which apply in terms of correct understanding (for example items, see Electronic Supplements 1 and 2). For 22 items, multiple choice answers were coded dichotomously (“0” = *wrong*, “1” = *correct*) because each incorrect selection of any of the 4–6 individual choice options indicated a fundamental misunderstanding of the information presented. One (.5) or two (.33 and .67) partial credits were defined for 29 and 16 items, respectively, because the answers can indicate correct understanding despite the partially incorrect response elements. Of the total 67 test items of the assessment instrument, 29 refer to the content area ECAP and 30 to COVID-19-IP. Eight test items refer to GHP topics for (expectant) parents. In 18 subsequent pretesting sessions (think aloud technique), we optimized the item bank regarding content validity, suitability, and comprehensibility (incl. optimization of distractor responses).

Data Collection and Study Sample

Data collection was conducted using an online questionnaire to obtain a sufficiently large and appropriate sample despite the limitations imposed by the COVID-19 pandemic. The data collection was divided into three parts (about 45 minutes each; interval: about 10 days). Part 1: Sociodemographic, allergy-, and health-related characteristics (e.g., self-efficacy, risk competence); Part 2: ECAP-HL items; Part 3: COVID-19-IP-HL items. Each participant was randomly assigned a test booklet containing 12–14 items. Details of the survey using test booklets are given in Electronic Supplement 3.

Data collection took place online from May 2021 to February 2022. The survey design and acquisition of participants were guided by the Tailored Design Method (TDM; Dillman et al., 2015). Participants were free to complete the HL items independently or as part of an individual digital video session. The Electronic Supplement 3 contains detailed information on the recruitment strategy. All 11 project staff members involved in recruitment had been trained prior to the start of recruitment and contacted participants in a standardized manner (e.g., providing identical information about study organization and participation).

Initially, we aimed to generate the sample entirely with parent couples. However, it turned out that the willingness of male partners to participate was limited. Hence, in the second half of the sampling period we decided to continue to recruit only women. Furthermore, because women with lower educational status were underrepresented, especially women without university entrance

qualifications were enrolled subsequently. Thus, we succeeded in recruiting a sufficiently large sample size of women that appropriately represents the German population regarding educational status. With a proportion of 53% of women with higher education entrance qualifications, the sample accurately reflects the school education situation in the overall female population in Germany in the age range 25–44 years (Bundesinstitut für Bevölkerungsforschung, 2021). For the final data analysis, we included only women: A mixed sample of couples and individuals would have implied that stochastic dependence within couples could only have been partially accounted for by hierarchical linear modeling and would have posed a particular problem because of the small number of participating male partners.

In the end, we included a total of 343 mothers (pregnant women: $n = 62$; 18%; mothers with children aged 0–3 years: $n = 281$; 82%) in the study sample; $n = 170$ (52%) are primipara. Their socioeconomic status as measured by the McArthur Scale (Hoebel et al., 2015) is slightly higher in the sample ($M = 5.87$; $SD = 1.42$) than in the reference standard sample of women aged 18–44 years ($M = 5.45$; $SD = 1.50$) (Hoebel et al., 2015). The average age of the respondents was 33.3 years (range: 20–47; $SD = 4.6$). Most of the women ($n = 193$, 57%) were affected by at least one allergy, whereas most of the children ($n = 291$, 84%) did not suffer from any allergy to date. The sample characteristics are shown in detail in Electronic Supplement 4.

Statistical Analysis

We conducted item and structural analyses using exploratory (EFA) and confirmatory factor analysis (CFA) for categorical, ordinal data, and IRT methods (restricted maximum likelihood [RML]; Harville, 1977) implemented in MPLUS 8.3. RML allows valid estimation in case of missing completely at random (MCAR), which was ensured because of the random assignment of test booklets by design. The comparison of models with different complexity (1-pl vs. 2-pl model) or dimensionality is based on measures from information theoretic measures (Rost & Langeheine, 1997), which allow an integrated assessment of model complexity and data fit (Akaike information criterion [AIC]; Bayesian information criterion [BIC]). We performed the final IRT analysis using Conquest software (Adams et al., 2020), using the multidimensional random coefficient multinomial logit (MRCML) approach, a multidimensional extension of the 1-pl or Rasch model. To test the fit of the items to the model assumption (logistic dependence of the solution probabilities on the latent trait), we calculated weighted and unweighted deviation measures (mean square deviations, MNSQ). The

model-consistent items had a value of about 1. Values higher than 1 (underfit) would indicate that responses to this item contain more stochastic information than the underlying model would predict; item fit values below 1 would indicate that the item fits better than predicted by the model (overfit). The latter usually occurs because of local dependencies of the items, which lead to higher item associations than expected based on latent trait expression. Thus, items with fit values < 1 do not pose a problem for the item validity as an indicator of latent trait and can be accepted as scale items (van der Linden, 2021). To avoid systematically penalizing larger sample sizes in IRT-based tests, items are usually not selected according to the criterion of significance. Instead, items with fit measures < 1.20 are considered acceptable (or 1.25 in the Pisa study; Frey et al., 2006). The quality of the overall test (group of all model-compliant items) is further assessed using the separation reliability (values $> .7$ acceptable; $> .8$ good; van der Linden, 2021). The person-specific estimates of the person parameters are obtained using the weighted-likelihood-estimate (WLE).

Results

Analysis of the Dimensional Structure Underlying the HL Items

EFA identified the single-factor structural assumption (1-DIM) as the best fitting for the pool of 67 items according to BIC (difference to the two-factor solution: $\Delta BIC = -271.70$; see Electronic Supplement 5; Burnham & Anderson, 2002). Confirmatory IRT analysis also revealed the most appropriate data fit for the 1-DIM model, which assumes only one latent dimension, Understand (BIC [1-pl] = 16,542.97). The AIC value, whose statistical validity is inferior to the BIC, showed a slightly better value for the 2-DIM-2pl model ($\Delta AIC = -7.59$). However, in the 2-DIM-2pl model, the very high latent correlation $r = .93$ of the two dimensions ECAP/GHP and COVID-19-IP also indicates the homogeneity of the item pool.

Thus, the unidimensional structure proves to be superior, although the items on COVID-19-IP not only refer to a different content area but were also measured at a different points in time (data collection at intervals of approximately 10 days on average). Accordingly, the construct intercorrelation of $r = .93$ in the 2-DIM-2pl model may also be interpreted in terms of construct stability or retest reliability over the test interval. In this respect, it is noticeable that the item group in the GHP domain should theoretically be similarly associated with the items on ECAP and COVID-19-IP. But the latent

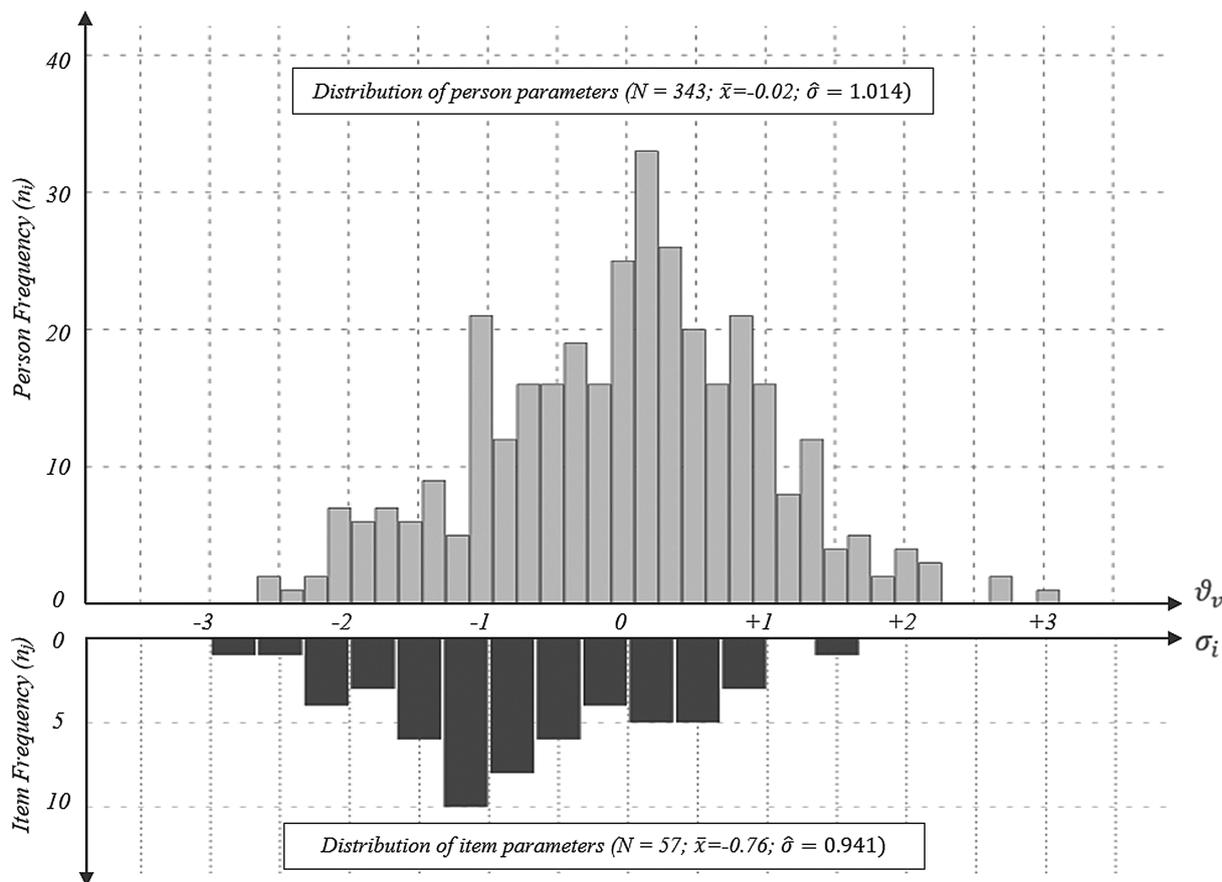


Figure 1. Wright map of person and item parameters (unidimensional 1-pl model).

correlation of GHP to the simultaneously measured content area ECAP is $r = .999$ and thus considerably higher than the latent correlation of GHP and COVID-19-IP: $r = .93$. Both the 1- and 2-factor model structures are further examined as plausible latent structural model alternatives.

Competing model definitions that distinguish the need to process number vs. number-independent information (numeracy; Kumar et al., 2010) or the presentation formats text, graphics, and tables (Nitsch et al., 2014) do not provide an adequate model alternative. In addition to the poorer values of the information-theoretic measures (see Electronic Supplement 5), the high latent construct correlations ($r = .993/.997$ and $r = .980-1.000$, respectively) indicate the psychometric identity of the assumed subfacets.

For all model definitions, BIC indicates the superiority of the more parsimonious 1-pl-/Rasch model over the 2-pl-/Birnbaum model (see Electronic Supplement 5). Thus the validity of analyzing the item pool according to the assumptions of the 1-pl model is warranted.

Analysis of Item and Scale Properties

The results of the analysis of the item pool using the MRCML estimation implemented in CONQUEST are shown in Table 1. In the first model estimations, the items A08, A09, A12, A30, C04, and C24 showed significant positive deviations from the expected value of 1 in the (un)weighted fit. Although all other items exhibited values below the critical value of 1.20 in the first estimation, the fit values of C09, C13, C21, and C19 rose beyond the critical value > 1.20 as a result of the stepwise item elimination. Finally, after excluding these 10 items, weighted fit values for the remaining 57 items did not deviate significantly upwards from 1 and did not exceed the critical value of 1.20. Unweighted fit values for a few single items were slightly above this limit (up to max. 1.25). For the final pool of 57 items, the one- and two-dimensional (ECAP/GHP vs. COVID-19-IP) models yield similar model fits ($\Delta AIC = 9.70$ and $\Delta BIC = 2.02$ each less than 10; Anderson & Burnham, 2002; see Electronic Supplement 5).

The 57 items conforming to the model assumptions have good separation reliability of .86, with a person variance on the latent trait of 0.95. The Wright map in Figure 1

Table 1. Item characteristics based on confirmatory 1pl-IRT-analysis for the 67 items assessing the HL facet Understanding Health Information (MCMRL estimation; CONQUEST)

Item	ECAP/GHP					COVID-19-IP					
	$h_i(\text{resp.})^a)$	$\sigma_i^b)$	$\lambda_i^c)$	$r_{it}^d)$	MNSQ	$h_i(\text{resp.})^a)$	$\sigma_i^b)$	$\lambda_i^c)$	$r_{it}^d)$	MNSQ ^{e)}	
	0/1/2/3				Unw. Weig.	0/1/2/3				Unw. Weig.	
A11	.13/.17/.70	-1.11	.54	.49	1.06 1.10	C05	.43/.33/.24	0.36	.43	.44	1.25 1.17
A06	.06/.25/.69	-1.49	.50	.45	1.12 1.03	C09 ^{g)}	.11/.55/.33				(1.26* 1.26*)
A41	.04/.20/.75	-1.80	.38	.41	1.21 1.00	C13 ^{g)}	.39/.28/.33				(1.30* 1.20*)
A17	.05/.56/.39	-1.16	.64	.52	0.90 0.91	C22	.37/.35/.29	0.12	.58	.56	0.96 0.98
A38	.23/.77	-1.28	.67	.51	0.82 0.87	C26	.25/.23/.38/.14	0.20	.58	.60	1.05 1.06
A14	.30/.36/.34	-0.01	.52	.56	1.03 1.05	C30	.48/.30/.23	0.48	.49	.48	1.12 1.04
A22	.28/.44/.28	0.07	.32	.37	1.23 1.21	C01	.30/.38/.32	-0.04	.51	.54	1.07 1.04
A33	.15/.85	-1.96	.64	.42	0.78 0.94	C14	.14/.86	-2.10	.09	.09	1.55 1.15
A37	.27/.73	-1.09	.38	.32	1.07 1.05	C20	.17/.23/.43/.16	-0.04	.60	.62	1.03 1.02
A30 ^{f)}	.14/.14/.32/.40				(1.29* 1.23*)	C24 ^{f)}	.27/.11/.28/.34				(1.56* 1.35*)
A36	.17/.22/.61	-0.94	.64	.63	0.81 0.95	C28	.16/.43/.41	-0.63	.39	.54	1.00 1.02
A27	.36/.64	-0.62	.60	.51	0.90 0.92	C10	.33/.31/.25/.11	0.55	.53	.60	1.05 1.04
A42	.31/.69	-1.06	.58	.49	0.89 0.92	C03	.36/.19/.35/.10	0.68	.58	.65	1.10 1.00
A02	.29/.71	-1.20	.75	.63	0.71 0.81	C15	.44/.27/.29	0.39	.65	.65	0.85 0.89
A09 ^{f)}	.12/.53/.28				(1.25* 1.29*)	C18	.13/.32/.55	-0.96	.65	.56	0.95 0.97
A18	.08/.27/.65	-1.58	.50	.47	1.08 1.04	C23	.33/.32/.36	0.01	.55	.56	1.08 1.04
A39	.14/.44/.42	-0.84	.59	.57	0.99 0.99	C27	.16/.53/.32	-0.42	.55	.53	0.98 0.99
A29	.26/.74/	-1.34	.52	.47	0.92 0.96	C02	.31/.36/.26	0.79	.59	.61	1.04 1.04
A07	.20/.27/.54	-0.83	.55	.56	1.06 1.08	C08	.07/.16/.27/.56	-0.90	.58	.57	1.04 1.04
A12 ^{f)}	.30/.52/.18				(1.28* 1.27*)	C11	.68/.25/.07	1.63	.67	.50	0.84 0.86
A24	.26/.74	-1.35	.71	.60	0.76 0.83	C17	.16/.16/.67	-1.07	.68	.62	0.85 1.00
A35	.26/.74	-1.35	.60	.51	1.02 0.90	C21 ^{g)}	.33/.29/.25/.13				(1.44* 1.30*)
A34	.19/.81	-1.83	.65	.51	0.83 0.90	C25	.23/.16/.61	-0.74	.72	.70	0.73 0.85
A04	.23/.38/.39	-0.35	.53	.57	1.01 1.01	C04 ^{f)}	.47/.35/.17/.02				(1.35* 1.25*)
A08 ^{f)}	.25/.52/.24				(1.29* 1.28*)	C07	.22/.78	-1.54	.14	.30	1.21 1.02
A15	.18/.30/.52	-0.79	.60	.60	0.94 0.97	C12	.28/.72	-1.19	.43	.37	0.96 0.98
A31	.23/.71	-1.10	.56	.49	0.91 0.96	C16	.19/.32/.50	-0.77	.44	.51	1.23 0.99
A19	.40/.60	-0.52	.54	.48	1.01 0.98	C19 ^{g)}	.30/.29/.32/.09				(1.33* 1.24*)
A21	.29/.27/.44	-0.33	.63	.62	0.89 0.94	C29	.44/.40/.16	0.61	.50	.50	1.05 1.02
G01	.07/.93	-2.80	.14	.10	1.24* 1.07	C06	.28/.23/.25/.25	0.13	.57	.64	1.07 1.08
G10	.17/.34/.49	-0.63	.59	.53	1.00 1.01						
G06	.12/.88	-2.26	.75	.51	0.63 0.89						
G09	.09/.91	-2.53	.79	.38	0.59 0.91						
G03	.14/.86	-2.28	.51	.38	0.94 0.98						
G07	.15/.85	-2.17	.42	.32	0.99 1.03						
G05	.14/.28/.58	-1.12	.57	.57	1.13 0.99						
G08	.65/.35	0.67	.34	.31	1.15 1.06						

Note. GHP = General Health Prevention, ECAP = Early Childhood Allergy Prevention, COVID-19-IP = COVID-19 infection prevention ^{a)} $h_i(\text{resp.})$ = relative frequencies of response categories: "0" = wrong, "1" = correct for dichotomous items; else "partial credit"; "2" = correct for three-level items; else "partial credit"; "3" = correct for four-level items; ^{b)} σ_i = item parameter (logit); ^{c)} λ_i = factor loading; ^{d)} r_{it} = corrected item-total correlation; ^{e)} MNSQ = mean square deviation (weighted/unweighted); ^{f)} weighted MNSQ > 1.20 in first estimation step (all items); ^{g)} weighted MNSQ > 1.21 (t-value = 2.0) in subsequent estimation steps after deletion of other misfitting items. * = Significant deviation of MNSQ from the expected value 1 ($p < .05$).

shows the distribution of the person and item parameters graphically. The estimated z-standardized person parameters (women's ability) are depicted at the upper

part of Figure 1. The distribution of the item parameters ($\bar{x}(\sigma) = -0.76$; range: -2.80-1.63, see bottom of Figure 1) is shifted to the negative relative to the distribution of the

Table 2. Correlations of IRT person-parameters in Understand Health Information with parents' sociodemographic variables as well as allergy- and COVID-19-specific person characteristics ($N = 343$)

Sociodemographic variables	
Age	.172***
Number of children	-.060
Age of the youngest child in months	.039
School graduation ^{a)}	.456***
Subjective socioeconomic state ^{b)}	.271***
Allergy-specific person characteristics	
Parents affected by allergy ^{c)}	.114*
Child affected by allergy ^{c)}	.054
Self-assessed impairment because of allergies ^{d)}	.127
COVID-19-specific person characteristics	
Already infected ^{c)}	.120*
Infection with severe progression in the personal environment ^{c)}	.089
I have regularly and extensively informed myself about the subject of COVID-19 ^{e)}	.120*
I felt burdened by COVID-19 ^{e)}	-.107*
I had a fear of a COVID-19 infection ^{e)}	-.030
It is important that everything possible be done to fight the COVID-19 pandemic ^{e)}	.129*

Note. ^{a)} Response categories: "0" = no university entrance qualification, "1" = University entrance qualification or higher; ^{b)} McArthur-Scale: 1 = low to 10 = high; ^{c)} response categories: "1" = no, "2" = yes; ^{d)} "Please indicate to what extent you feel impaired by your allergy(s) or allergic disease(s) in your everyday life in general?", response categories: "1" = very little/not at all to "6" = very strong; ^{e)} high values = high agreement. * $p < .05$, ** $p < .01$, *** $p < .001$.

person parameters ($\bar{x}(g) = -0.02$ range: -2.64–3.16). Thus, the item pool is rather "easy" regarding the ability of the participating women. Accordingly, the solution probabilities for the items tend to be above 50 %.

Association of Person Parameters on the Trait Understand Health Information with Further Women's Characteristics

Person-parameter estimates of women's capability to Understand Health Information are positively related to school graduation ($r = .46$) and self-assessed socioeconomic status ($r = .27$; Table 2). The person parameter is weakly correlated with women's allergy status ($r = .11$), occurred COVID-19 infection ($r = .12$), informedness about COVID-19 ($r = .12$), perceived burden by the pandemic ($r = -.11$), and the statement that everything should be done to fight the pandemic ($r = .13$).

Discussion

We developed and psychometrically validated a content-valid unidimensional item pool for the HL facet Understand Health Information in a population of pregnant women and mothers of infants. All 67 items were signi-

ficantly associated with the first principal component and exhibited significant item-total correlations. Only 10 items showed significant deviations from the response structure as expected by the unidimensional 1-pl-/Rasch model (Research question 3). The final 57 items reflect the information of the latent dimension according to the assumptions of the 1-pl-/Rasch model (van der Linden, 2021) (Research question 1). Although the items were developed separately for the two application domains ECAP and COVID-19-IP, the content domains do not map as separable facets or dimensions (Research question 2). This provides a promising assessment anchor for operationalizing and critically examining the HL facet Understand Health Information as a generic component, which is basal in any HL model (Sørensen et al., 2012). As expected, the assessed HL facet Understand Health Information is substantially associated with educational qualification ($r = .46$) and subjective socioeconomic status ($r = .27$; Research question 4).

Thus, an innovative assessment is available to overcome the shortcomings of procedures based on subjective self-assessments (Schmuckle & Egloff, 2011). There has been a strong call for the development and use of assessment instruments that meet IRT criteria to take advantage of the beneficial psychometric properties of IRT-based scales (Cella et al., 2002). IRT models have already been successfully applied for HL self-assessment scales (Sijtsma & van der Ark, 2021). The present paper shows that the

established standards in educational research for objective competency scales based on performance items (Rupp et al., 2010) can also be adopted to assess distinct HL facets. A calibrated item pool can now be used to characterize the ability of pregnant women and mothers of infants to Understand Health Information in terms of clearly operationalized and psychometrically validated construct and task definitions (van der Linden, 2021).

This provides a promising perspective for the explicit definition of critical HL levels in terms of defined task requirements (competence level modeling; Rupp et al., 2010). These levels can serve as diagnostic criteria as well as learning and development goals (Wilson, 2005). By definition, valid level-specific subtests can be built from the item pool, or adaptive test procedures can be developed that are optimally adjusted to women's ability levels. This avoids the regularly reduced measurement accuracy and limited sensitivity to change of static tests (van der Linden, 2021). IRT scales are generally associated with manifold additional benefits in diagnostic and evaluative applications (Wirtz & Böcker, 2007).

Last but not least, the measurement properties of the item pool may be continuously critically examined and enhanced. It was empirically proven for the two domains ECAP and COVID-19-IP that the construct facet Understand Health Information can be considered generic. With the psychometric properties now known and the relationship of the items to the underlying latent trait identified, items or item groups from other application domains (e.g., HL related to health-conscious parental lifestyle) for other populations (e.g., adults without children) or various information formats can be tested at a much more advanced level. The potential of mixed-Rasch modeling (Wirtz et al., 2005), differential item functioning (Wirtz & Farin, 2018), and linking methods (Kolen & Brennan, 2014) open up a broad spectrum for analyzing the fairness and generalizability of measurement properties. All of these methods serve in particular to enhance content, differential validity, and construct validity (Messick, 1995). Since research on HL in public health usually relies on generic operationalizations to make valid comparisons across populations and domains, the IRT-conform item bank now provides a psychometrically high-quality basis to better substantiate these claims empirically (Cella et al., 2010).

A particular challenge of the study was that the performance assessment was conducted in the context of a questionnaire study lasting approximately 135 minutes, which had to be conducted entirely online because of the COVID-19 pandemic. Nevertheless, the women's motivation to participate was high until the end of the third part of the survey. The participating women rated the study content as interesting ($M = 1.60$, $SD = 0.59$) and would recommend participation to other parents ($M = 1.70$,

$SD = 0.75$) ("1" = *strongly agree* to "4" = *strongly disagree*). The following aspects may be regarded as essential for this successful online-based distance testing: (1) The women took part in a study whose content concerned their child's healthcare, an essential aspect of their life situation and their parental responsibility. (2) An individual contact person accompanied study participants throughout the survey process. They could use the option to be accompanied by a video conference. (3) The performance test was not dominantly perceived as an examination situation. The items (see Electronic Supplements 1 and 2) consisted mainly of knowledgeable and interesting health-related learning content that directly relates to the reality of parental life. The acquisition of new knowledge was dominant, and the subsequent check – i.e., the actual performance assessment – was perceived more as a self-assurance of correct individual understanding. (4) Frustration in the course of processing was also avoided by the items not being perceived as too difficult. The item pool assesses women's literacy levels most accurately in the lower ability range (see Figure 1; the average rate of correct item solutions was over 50%). Although this proved beneficial for women's motivation to solve the individual items, further assessment development should strive for a homogeneous distribution of item difficulties across the entire ability spectrum. (5) The women also received 30 EUR as an expense allowance for participating in all three survey terms. This amount was rated on a 4-point scale between "1" = *totally appropriate* to "2" = *rather appropriate* with $M = 1.69$ ($SD = 0.72$).

A limitation of the study is that online distance testing is not standardized and controlled to the same extent as face-to-face testing. Thus, the influences of uncontrolled disturbances during processing cannot be excluded. This problem could be reduced by enabling women to give feedback during processing. The electronically documented processing times indicate serious testing. Furthermore, the representativeness of the sample is not ensured, especially since participation was voluntary and an incentive was paid. Because of the low participation rate of men ($N = 178$), it was impossible to analyze the data for the participating couples using hierarchical linear models. Because of the test-booklet design, each individual item could be completed by only about $N = 60$ men. A sample size of $N = 150$ would have been necessary to ensure stable IRT estimation results (van der Linden, 2021). Future research could conduct a structural analysis for fathers and mothers together, not only to assess them comparatively but also to consider dependencies within the couple constellation regarding parental HL.

Because the study sample exhibits an average level of education (53% university entrance qualification), we assume good generalizability of the findings. Expected

biases because people in less educated social strata tend to participate in studies on HL less probably (Sørensen et al., 2012) and tend to exhibit lower HL levels could thus be avoided. Hence, follow-up surveys are currently being conducted or planned for parents with low educational qualifications and an immigrant background (Arabic translation version). Furthermore, by using 1-pl IRT-based models, it could be ensured as best as possible that the estimation and analysis of item characteristics are not significantly influenced by this specific sample characteristic (criterion of specific objectivity; van der Linden, 2021).

In summary, we developed and validated a high-quality operationalization of the HL facet Understand Health Information in terms of content and psychometrics. The 57 items validly represent the fundamental HL facet Understand Health Information, despite the different content of the health (prevention) domains – ECAP, GHP, COVID-19-IP. The fit of the items to the 1-pl-/Rasch model offers a psychometrically excellent basis for comparative studies of the construct and item properties (van der Linden, 2021). The conceptual anchoring in existing HL models (Nutbeam, 2000; Soellner et al., 2010; Sørensen et al., 2012) enables the integrated analysis (especially structural equation models and nomological networks) of the construct facet Understand Health Information in more complex, multilevel process models, regarding the importance of and interaction with antecedent, consequent, and moderating HL components and lifeworld framework conditions.

Electronic Supplementary Material

The electronic supplementary material is available with the online version of the article at <https://doi.org/10.1026/0012-1924/a000298>

ESM 1. Test booklet with items on ECAP and GHP.

ESM 2. Test booklet with items on COVID-19-IP.

ESM 3. Data collection using a test booklet design/recruitment strategy.

ESM 4. Characteristics of the study sample.

ESM 5. Exploratory and confirmatory analysis of the 67 items on Understand Health Information.

References

Adams, R. J., Wu, M. L., Cloney, D., & Wilson, M. R. (2020). *ACER ConQuest: Generalised item response modelling software [Computer software], Version 5*. Australian Council for Educational Research.

- Andrulis, D.P., & Brach C. (2007). Integrating literacy, culture, and language to improve healthcare quality for diverse populations. *American Journal of Health Behavior*, 31(Suppl 1), 122–133. <https://doi.org/10.5555/ajhb.2007.31.supp.S122>
- Borges, K., Sibbald, C., Hussain-Shamsy, N., Vasilevska-Ristovska, J., Banh, T., Patel, V. et al. (2017). Parental Health Literacy and Outcomes of Childhood Nephrotic Syndrome. *Pediatrics*, 139(3), e20161961. <https://doi.org/10.1542/peds.2016-1961>
- Bundesinstitut für Bevölkerungsforschung. (2021). *Bund-Länder Demografieportal: Schulabschluss der Bevölkerung nach Alter und Geschlecht, 2020* [Federal-State Demography Portal: Graduation of the population by age and gender]. Retrieved from www.demografie-portal.de/DE/Fakten/schulabschluss.html
- Burnham, K. B., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer.
- Cella, D., Chang, C. H., & Heinemann, A. W. (2002). Item response theory (IRT): Applications in quality of life measurement, analysis and interpretation. In M. Mesbah, B. F. Cole, & M. T. Lee (Eds.), *Statistical methods for quality of life studies* (pp. 169–185). Springer. https://doi.org/10.1007/978-1-4757-3625-0_14
- Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., Amtmann, D., Bode, R., Buysse, D., Choi, S., Cook, K., Develis, R., DeWalt, D., Fries, J. F., Gershon, R., Hahn, E. A., Lai, J., Pilkonis, P., Revivki, D., Rose, M., Weinfurt, K., Hays, R., & PROMIS Cooperative Group (2010). Initial item banks and first wave testing of the patient-reported outcomes measurement information system (PROMIS) network: 2005–2008. *Journal of Clinical Epidemiology*, 63, 1179–1194. <https://doi.org/10.1016/j.jclinepi.2010.04.011>
- Cluver, L., Lachman, J. M., Sherr, L., Wessels, I., Krug, E., Rakotomalala, S., Blight, S., Hillis, S., Bachman, G., Green, O., Butchart, A., Tomlinson, M., Ward, C. L., Doubt, J., & McDonald, K. (2020). Parenting in a time of COVID-19. *The Lancet*, 395(10,231), e64. [https://doi.org/10.1016/S0140-6736\(20\)30736-4](https://doi.org/10.1016/S0140-6736(20)30736-4)
- DeWalt, D. A., & Hink, A. (2009). Health literacy and child health outcomes: A systematic review of the literature. *Pediatrics*, 124(3), 265–274. <https://doi.org/10.1542/peds.2009-1162B>
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2015). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.
- Dresch, C., Schulz, A. A., & Wirtz, M. A. (2021). Modellierung und Messung elterlicher Gesundheitskompetenz im Bereich frühkindlicher Allergieprävention [Modeling and measuring parental health literacy in early childhood allergy prevention]. In K. Rathmann, K. Dadaczynski, O. Okan, & M. Messer (Eds.), *Gesundheitskompetenz* (pp. 1–11). Springer. https://doi.org/10.1007/978-3-662-62800-3_139-1
- Eid, M., & Schmidt, K. (2014). *Testtheorie und Testkonstruktion* [Test theory and test construction]. Hogrefe.
- Frey, A., Taskinen, P., Schütte, K., Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (2006). *PISA 2006 Skalenhandbuch* [PISA 2006 scale manual]. Waxmann.
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358), 320–338. <https://doi.org/10.2307/2286796>
- Haun, J. N., Valerio, M. A., McCormack, L. A., Sørensen, K., & Paasche-Orlow, M. K. (2014). Health literacy measurement: An inventory and descriptive summary of 51 instruments. *Journal of Health Communication*, 19(2), 302–333. <https://doi.org/10.1080/10810730.2014.936571>
- Hoebel, J., Müters, S., Kuntz, B., Lange, C., Lampert, T. (2015). Messung des subjektiven sozialen Status in der Gesundheitsforschung mit einer deutschen Version der MacArthur Scale

- [Measuring subjective social status in health research with a German version of the MacArthur Scale]. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 58, 749–757. <https://doi.org/10.1007/s00103-015-2166-x>
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts: State of the art and future prospects* (pp. 3–22). Hogrefe.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices. Statistics for social and behavioral sciences* (3rd ed.). Springer.
- Kumar, D., Sanders, L., Perrin, E. M., Lokker, N., Patterson, B., Gunn, V., Finkle, J., Franco, V., Choi, L., & Rothman, R. L. (2010). Parental understanding of infant health information: Health literacy, numeracy, and the Parental Health Literacy Activities Test (PHLAT). *Academic Pediatrics*, 10(5), 309–316. <https://doi.org/10.1016/j.acap.2010.06.007>
- Mackert, M., Champlin, S., Su, Z., & Guadagno, M. (2015). The many health literacies: Advancing research or fragmentation? *Health Communication*, 30(12), 1161–1165. <https://doi.org/10.1080/10410236.2015.1037422>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- Nitsch, R., Fredebohm, A., Bruder, R., Kelava, A., Nacarella, D., Leuders, T., & Wirtz, M. A. (2014). Students' competencies in working with functions in secondary mathematics education: Empirical examination of a competence structure model. *International Journal of Science and Mathematics Education*, 13(3), 657–682. <https://doi.org/10.1007/s10763-013-9496-7>
- Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int*, 15(3), 259–267. <https://doi.org/10.1093/heapro/15.3.259>
- Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults. *Journal of General Internal Medicine*, 10, 537–541. <https://doi.org/10.1007/BF02640361>
- Pohontsch, N., & Meyer, T. (2015). Das kognitive Interview: Ein Instrument zur Entwicklung und Validierung von Erhebungsinstrumenten [Cognitive interviewing: A tool to develop and validate questionnaires]. *Die Rehabilitation*, 54(1), 53–59. <https://doi.org/10.1055/s-0034-1394443>
- Rost, J., & Langeheine, R. (Eds.). (1997). *Applications of latent trait and latent class models in the social sciences*. Waxmann.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford.
- Schmuckle, S. C., & Egloff, B. (2011). Indirekte Verfahren zur Erfassung von Persönlichkeit ("Objektive Persönlichkeitstests") [indirect procedures for the assessment of personality („objective personality tests“)]. In L. F. Hornke et al. (Eds.), *Enzyklopädie der Psychologie: Bd. 4: Themenbereich B, Serie 2, Psychologische Diagnostik* (pp. 73–120). Hogrefe.
- Sijsma, K., & van der Ark, L. A. (2021). Advances in nonparametric item response theory for scale construction in quality-of-life research. *Quality of Life Research*, 31, 1–9. <https://doi.org/10.1007/s11136-021-03022-w>
- Soellner, R., Huber, S., Lenartz, N., & Rudinger, G. (2010). Facetten der Gesundheitskompetenz: Eine Expertenbefragung. Projekt Gesundheitskompetenz [Facets of health literacy: An expert survey. Health literacy project]. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung: Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (pp. 104–144). Beltz.
- Soellner R., & Rudinger, G. (2018). Gesundheitskompetenz [Health literacy]. In C.-W. Kohlmann, C. Salewski & M.A. Wirtz (Eds.), *Psychologie in der Gesundheitsförderung* [Psychology in health promotion] (pp. 59–72). Hogrefe.
- Sørensen, K., van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J. M., Slonska, Z., Brand, H., & (HLS-EU) Consortium Health Literacy Project European (2012). Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, 12(1), 80. <https://doi.org/10.1186/1471-2458-12-80>
- Sørensen, K., van den Broucke, S., Pelikan, J. M., Fullam, J., Doyle, G., Slonska, Z., Kondilis, B., Stoffels, V., Osborne, R. H., Brand, H., & HLS-EU Consortium (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), 948. <https://doi.org/10.1186/1471-2458-13-948>
- Van der Linden, W. J. (Ed.). (2021) *Handbook of item response theory*. Chapman & Hall.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber Publishers.
- Wilson, M. (2005). *Constructing measures. An item response modeling approach*. Erlbaum.
- Wirtz, M., & Böcker, M. (2007). Das Rasch-Modell – Eigenschaften und Nutzen für die diagnostische Praxis [The Rasch Model – Properties and benefits for diagnostic practice]. *Die Rehabilitation*, 46(4), 238–245. <https://doi.org/10.1002/pon.3092>
- Wirtz, M. A., & Farin, E. (2018). Generische und indikationsspezifische Messeigenschaften des IRES-24-Patientenfragebogen Ein Vergleich der Skalenstruktur bei orthopädischen und neurologischen Rehabilitationspatientinnen und -patienten mittels Differential Item Functioning [Generic and indication-specific measurement properties of the IRES-24 patient questionnaire. A comparison of the scale structure in orthopedic and neurological rehabilitation patients using Differential Item Functioning]. *Diagnostica*, 64(2), 74–83. <https://doi.org/10.1026/0012-1924/a000193>
- Wirtz, M. A., Dresch, C. J., & Schulz, A. A. (2021). Structural modelling and assessment of health literacy in allergy prevention of new parents by means of item-response-theory. *Psych-Archives*. <https://doi.org/10.23668/psycharchives.4551>
- Wirtz, M., Farin, E., Bengel, J., Jäckel, W., Hämmerer, D., & Gerdes, K. (2005). IRES-24 Patientenfragebogen – Entwicklung der Kurzform eines Assessmentinstrumentes in der Rehabilitation mittels der Mixed-Rasch-Analyse [IRES-24 patient questionnaire – development of the short form of an assessment instrument in rehabilitation using mixed Rasch analysis]. *Diagnostica*, 51(2), 75–87. <https://doi.org/10.1026/0012-1924.51.2.75>

Conflict of Interest

The authors declare that there are no conflicts of interest.

Publication Ethics

The German Psychological Society has given a positive ethical vote for this study (registration no. MAW 112018).

Authorship

Dr. Anja Alexandra Schulz and Dr. Carolin Dresch share the first authorship.

Funding

The study was conducted within the project “Structural Modelling and Assessment of Health Literacy in Allergy Prevention of New Parents” funded by the German Research Foundation (Deutsche Forschungsgemeinschaft – DFG; G.Z.: WI-3210/7 – 1). This project is part of the DFG Research Group “Health Literacy in Early Childhood Allergy Prevention: Parental Competencies and Public Health Context in a Shifting Evidence Landscape” [FOR 2959; G.Z.: AP 235/3 – 1].

Open Access publication enabled by the University of Education Freiburg.

ORCID

Anja Alexandra Schulz

 <https://orcid.org/0000-0002-9591-5689>

Carolin Dresch

 <https://orcid.org/0000-0001-5593-2954>

Andrea Heiberger

 <https://orcid.org/0000-0002-2446-9626>

Markus Antonius Wirtz

 <https://orcid.org/0000-0002-9392-0484>

Dr. Anja Alexandra Schulz

University of Education Freiburg

Kunzenweg 21

79117 Freiburg

Germany

anja.schulz@ph-freiburg.de