

**Structural modelling and assessment of  
health literacy in allergy prevention of new parents  
by means of Item-Response-Theory**

## **Abstract**

Parents of newborn children need appropriate early childhood allergy prevention specific health literacy (ECAP-HL) to act responsible and adjuvant concerning the child's allergy or allergy disposition, respectively. For diagnostic and evaluative assessment purposes there is a need to define the construct ECAP-HL and its structural facets. A comprehensive multidimensional assessment instrument will be developed, which allows measuring and analyzing parents' facet levels of ECAP-HL.

Health literate people are able to find and process health related information and to decide and act purposively to achieve beneficial health goals. ECAP-HL is a more specific construct addressing parent's perspective on their child's allergy. ECAP-HL is closely related to allergy specific knowledge, skills and capacities. Parents must be able to assess the child's allergy related health state and risk factors, to pose answerable questions, to search relevant health information, to critically appraise evidence, to evaluate, and to apply the resulting conclusions in allergy prevention behavior. When modeling and evaluating ECAP-HL, it is important to consider that parents' health-related attitudes, beliefs, and behaviors regarding the newborn's allergy-specific health situation are influenced by the multiple demands of an extraordinary life situation. A comprehensive literature search on ECAP and HL as well as interviews with new parents and experts on ECAP-HL will serve as basis for the conceptual and empirical research. To ensure a valid assessment, scenarios will be defined reflecting everyday life situations which are typical in the life of new parents. A comprehensive item pool will be developed and answered by N = 600 new and expectant parents as well as N = 600 young adults without children. Multidimensional construct properties and facet-specific item scales will be analyzed using Item-Response models. The finalized assessment will be validated in a second data collection phase (N = 300 parents, N = 300 young adults without children).

**Keywords:** Health literacy, allergy prevention, competence modelling, Item-Response-Theory, new parents

## 1. Introduction

This study will be conducted within a DFG-founded research group which investigates health literacy (HL) within early childhood allergy prevention (ECAP). By reducing the evidence-practice gap concerning ECAP, the HL of new parents is aimed to be empowered by a user-friendly design and organization of the health system as well as health information in general. Higher HL for all parents is expected to decrease the risk of allergies among their children. For diagnostic and evaluative assessment purposes there is a need to define the construct *early childhood allergy prevention specific health literacy (ECAP-HL)* and its structural facets. Based on this conceptual foundation a comprehensive multidimensional assessment instrument will be developed, which allows measuring parents' facet levels of ECAP-HL.

The construct of ECAP-HL will be investigated according to existing models of the more general construct *health literacy (HL)*. HL can be defined as the ability to purposefully *access*, *comprehend*, and *reflect* health related information as well as to *apply* the resulting conclusions to problems with respect to health-related decisions to maintain or improve quality of life during the life course (1, 2). Health literate people are able to find and process health related information and to decide and act purposively to achieve beneficial health goals. This requires people not only to be able to understand health related numerical information and texts but also to act self-reliant and responsible based on appropriate health information and valid health concepts (3-5). Peoples' individual health related capacities, skills and competencies are decisive to process health information appropriately and to act goal-oriented in a complex health system.

Several model conceptions of HL have been proposed and analyzed. Generally, HL is conceptualized as a multidimensional construct comprising different literacy facets like *Information Literacy*, *Statistical Literacy*, and *Critical Thinking/Evidence based reasoning* (6, 7). *Information literacy* is defined as the ability to formulate appropriate questions and to search and evaluate relevant information. *Statistical literacy* refers to the ability to read and organize data, and interact with different data representations. Statistical literate people understand numerical information validly. *Critical thinking or Evidence based reasoning* denotes peoples' ability to critically appraise information and draw appropriate conclusions from health information.

Focusing on decision processes these literacy facets are assumed to be decisive to process health-related information appropriately: people should be able (i) to *assess* their health state and *identify* health related needs, (ii) to *pose answerable questions* for a specified health problem, (iii) to *search* for relevant information, (iv) to *critically appraise* evidence according to their individual health situation, (v) to *evaluate* and (vi) to *apply* the resulting conclusions. Comparable process descriptions can be found in models of evidence-based processes to

construct new knowledge in general (8, 9). *Functional* (health related knowledge; basic health related skills), *perceptual* and *emotional* (accountability, self-perception) as well as *action-oriented components* (handling health-related information; self-control; self-regulation; communication and cooperation) are assumed to be essential (10-12).

Nutbeam (13) proposed three different types of HL: (i) *Functional HL* refers to basic skills like reading of health-related texts, numbers or symbolic information. (ii) *Interactive HL* denotes more advanced cognitive and literacy skills, which are important to actively participate in everyday situations, extract information and derive meaning from different forms of communication. Interactive HL is important to use and apply health-related information flexibly in different situations and varying circumstances. (iii) *Critical HL* denotes advanced cognitive skills which allow to analyze information critically, and to make literate decisions to enhance a self-determined life, control over life events and situations. Critical HL is associated with knowledge and skills fostering individual autonomy, empowerment, valid health related decision-making, self-management competencies as well as social determinants of health (14) and appropriate media use (15).

From a life-span developmental view Mancuso (16) emphasizes *capacity*, *comprehension* and *communication* as main aspect of HL. *Capacity skills* include aspects as gathering, analyzing, and evaluating health information for credibility and quality. Furthermore, goal-oriented decision making and acting in the health care system are subsumed.

*Comprehension skills* enable people to understand and appraise complex health information like patient education materials or medication directions. *Communication skills* are necessary to interact with other people, to spell individual demands and to react in interpersonal interactions.

The concept *public health literacy* (17) specifies necessary information (formats), skills and competences enabling people to understand, decide, act and participate in public health contexts: (i) *Conceptual foundations*: Basic knowledge and information enabling people to understand and to take action; individuals and groups should be able to reflect and discuss core public health issues and problems. (ii) *Critical skills* are required to obtain, to process and to evaluate public health related information, and to act upon this information in order to foster benefits on personal as well as community level. (iii) *Civic orientation* denotes skills and resources necessary to appraise societal health concerns and to be engaged in health-related civic discourses.

Sorensen et al. (1) proposed an integrative and the most general definition of HL based on existing models. HL is characterized as an “asset for improving people's empowerment within the domains of healthcare, disease prevention and health promotion.” Each domain includes the four components *access*, *understand*, *appraise* and *apply* health related information.

Accordingly, in the *domain of health care* parents should be able to access information on medical issues, to understand medical information and to derive meaning, to interpret and evaluate medical information and to make informed decisions and comply with medical advices concerning the medical treatment of allergies. Health literate *disease prevention* requires parents to access, to interpret and evaluate information of allergy risk factors and to make informed decisions on allergy risk factors. In the *domain of health promotion* updating on determinants of health in the social and physical environment and adapting health promotion behavior accordingly are assumed to be crucial.

Additionally, the properties of the environment and the health care system have to be regarded: Individual capacities and skills lead to more health literate behaviors in a patient-oriented complex health-care system (person-system-interaction). Accordingly, health literate people should be able to apply health concepts and information to novel situations, and to participate in public and private dialogues about health, medicine, scientific knowledge, and cultural beliefs (18).

The construct HL has to be investigated in nomological networks considering antecedent, mediating, moderating aspects as well as its consequences. Distal factors (e. g. societal and environmental determinants) as well as proximal factors (e. g. personal determinants as age, gender or literacy, and situational determinants as social support or media use) may influence individuals HL decisively (19). Operational, interactive, autonomous, informational, contextual, and cultural competencies are considered as necessary antecedents of HL (16). Although motivational and environmental aspects are important to validly predict health-related behavior, these aspects are typically regarded as influential or moderating variables, which are not core facets of HL (11, 20).

Several instruments have been developed assessing general HL based on objective performance data (e. g., REALM, TOFHLA, BRIEF, NVS) or on subjective judgements (e. g., HeLMS (21); HLS-EU-Q (22)). The *Rapid Estimate of Adult Literacy in Medicine (REALM*, (23)) consists of 66 or 125 health related items. The test score equals the number of words people are able to pronounce correctly. From these scores corresponding health literacy levels can be determined. The ability to comprehend printed health materials, numeracy, and information seeking/navigation are not covered by the REALM items (24). The *test of functional health literacy in adults (TOFHLA)* assesses individuals' ability to read and understand health-related information (50 items). Omitted words have to be selected from four choices in each case. The *Medical Term Recognition Test (METER*; recognition of medical words (25)), the *Functional Health Literacy Test (FHLTs*; (26)) assess functional health literacy levels of the general public as well as of patients in healthcare settings.

In the light of HL construct definitions, reading skills or correct pronouncing (REALM) or choosing words (TOFHLA) is at most an indirect and basic indicator of health related information seeking, thinking, deciding and acting. Accordingly, despite the frequent use of these two instruments, several studies documented considerable conceptual or psychometric shortcomings (27, 28) or even do not recommend their application to assess HL comprehensively (24). Hence, assessments of HL and ECAP-HL, respectively, should measure all facets proposed in a comprehensive structural model of HL (particularly, *access*, *understand*, *appraise* and *apply*). Test contents should validly address information formats, decisional situations and behavior patterns which are typical in everyday life of new parents.

### *Psychometric standards in competence structure and competence level modelling*

Modern standards of *Item-Response-Theory (IRT)* have to be adopted to ensure appropriate modelling of competence/literacy structures and psychometric sound competence/literacy assessments (29-31). Item-Response-Theory assumes that the probability of a correct response to an item depends logistically on a continuous latent trait. Latent traits are defined as dimensions reflecting persons' literacy facets. Unidimensional models assume that only one general latent trait determines the response probability to each item within the item pool. Multidimensional item models consist of more than one latent trait. For multidimensional competence assessments the response probability to each item should be determined by one single latent trait (between-item-multidimensionality). Hence, each latent trait is reflected by a unique and distinct item group. The responses to each item group are used to estimate metric person literacy levels (person parameters). Persons' metric literacy value may be assigned to ordinal literacy levels.

Generally, structure models and assessment scale meeting the assumptions of IRT ensure beneficial psychometric properties, because the mathematical relation between person answers and literacy levels is explicitly defined. Item-Fit-statistics can be adopted to estimate and test the validity of the model assumptions. Additionally, overall model-tests allow for testing the model assumptions for the whole item pool. Hence, construct structure, unidimensionality of multi-item scales as well as item-scale-associations can be tested empirically. IRT-conform scales are unidimensional by definition and metric literacy estimates can be determined. Potential violations of item-fairness, intergroup/cultural fairness and scale homogeneity can be analyzed by sophisticated psychometric analyses.

IRT modelling has proved to be advantageous in several fields of competence and literacy modelling. E. g. for the Educational Research Literacy (ERL) valid multifaceted competence models and assessment has been identified. This is of special relevance for the research on HL, because the underlying model components of ERL are similar as those assumed for HL (1): 1) *ask* an adequate question that concerns a problem, 2) *search* for relevant information

(e.g., via databases), 3) *evaluate/appraise* evidence, and 4) *integrate information* to draw conclusions (32, 33).

### *Research objectives and research hypotheses*

Starting from existing models and items measuring HL the specific properties of ECAP-HL have to be identified. ECAP-HL is a construct in a specific population (new parents) that is related to specific health aspects (allergy). In existing HL assessment general HL facets of the tested subjects are measured. Modelling and assessing ECAP-HL has to take into account that parents' health related attitudes; beliefs and behavior concerning the newborns' allergy-specific health situation are affected by the versatile and complex demands of a very challenging and exceptional life situation. Hence, parents' ECAP-HL is affected by the specific demands of this life period and may be affected by manifold changes concerning several aspects and life conditions of new parents.

Appropriate models and assessments of ECAP-HL should allow to measure different facets of ECAP-HL (e. g. access, understand, appraise, and apply). Furthermore, personal and situational determinants have to be regarded. To ensure a valid assessment item contents should comprise test scenarios reflecting everyday situations and tasks of new parents. A valid structure model of ECAP-HL has to be tested confirmatory based on the assumptions of IRT. Unidimensional assessment scales and test-fair item groups which measure the ECAP-HL will be developed. Generally, developing a reliable as well as a content and construct valid instrument is fundamental to assess and evaluate the different facets of ECAP-HL appropriately.

We aim at developing and validating a comprehensive multi-facet structural model of ECAP-HL and a multidimensional assessment instrument measuring germane facets of ECAP-HL for new parents. WP 8 is a cross-sectional project in the HELICAP-network, because a clarified, explicated and shared understanding of ECAP-HL provides an important precondition for concertedly investigating and communicating about health literacy in early childhood allergy prevention.

The sample we include both new or expectant parents (N = 600) and adults without children (N= 600). All objectives hypotheses will be tested for both groups separately and in the whole group.

The specific objectives of the study will be:

1. Starting from existing models and assessments of HL general and specific aspects of ECAP-HL will be determined on a conceptual level.
2. Facet-specific item pools will be defined and analyzed.

3. Structural, diagnostic and evaluative characteristics of the item pools will be evaluated empirically.
4. A comprehensive assessment instrument will be defined and validated (content, construct and criterion validity; sensitivity to change). Subgroup-specific (e. g. age-groups, gender, allergy status of parents, health status of the child) assessment properties will be analyzed on item- as well as on facet-level.

The central hypotheses are:

- Hypothesis 1: For each ECAP-HL-component *search*, *understand*, *appraise* and *act* a unidimensional item-pool meeting the assumptions of Item-Response-Theory.
- Hypothesis 2: For each ECAP-HL-component the 1-pl-model (Rasch-model) allows for a better data-fit than the 2-pl-model (Birnbaum-model).
- Hypothesis 3: The entire item-pool assessing the four components *search*, *understand*, *appraise* and *act* meets the assumptions of a four-dimensional Item-Response-model (between-item-multidimensionality).
- Hypothesis 4: The 1-pl-model (Rasch-model) allows for a better data-fit in the four-dimensional model than the 2-pl-model (Birnbaum-model).
- Hypothesis 5: The level in each ECAP-HL component is associated by
- (a) Gender (male vs. female): Women exhibit higher ECAP-HL.
  - (b) Age: Older adults exhibit higher ECAP-HL.
  - (c) Socio-economic status: Higher socio-economic status is associated with higher ECAP-HL.
  - (d) Allergy status: Adults suffering from allergy exhibit higher ECAP-HL.
  - (e) Experience of threat: Higher experience of allergy specific threat is associated with higher ECAP-HL.
  - (f) Allergy-specific knowledge: Higher Knowledge is associated with higher ECAP-HL.
  - (g) Control beliefs: Higher control beliefs are associated with higher ECAP-HL.
  - (h) Self-efficacy: Higher self-efficacy is associated with higher ECAP-HL.
  - (i) Prevention motivation: Higher Prevention motivation is associated with higher ECAP-HL.
  - (j) Risk competence: Higher risk competence is associated with higher ECAP-HL.



- (k) Parenthood: New or expectant parents exhibit higher ECAP-HL than adults without children.
- (l) First child vs. not-first child: Parents who give birth to the first child have lower ECAP-HL.

Hypothesis 5: Scale structures are moderated (Differential-Item-Functioning) by parent's characteristics:

- (a) Gender (male vs. female)
- (b) Age
- (c) Socio-economic status
- (d) Allergy status
- (e) Experience of Threat
- (f) Allergy-specific knowledge
- (g) Control beliefs
- (h) Self-efficacy
- (i) Prevention motivation
- (j) Risk competence
- (k) Parenthood
- (l) First child vs. not-first child

## **2. Materials and Methods**

### *Task Developing a conceptual model ECAP-HL*

A comprehensive literature review will be conducted to identify construct facets and definitions of HL and to identify special aspects of ECAP-HL in new parents. All researchers of the HELICAP researchers will take part in a thematic focused. A moderated discussion applying metaplan technique and scenario based working groups will be applied to identify all relevant facets of ECAP-HL from the perspectives of each HELICAP project. Based on this a fundamental model conception will be defined (Delphi-study; N = 10).

Additionally, guided interviews will be conducted with N = 20 new parents. Allergy and health related views, attitudes, beliefs and experiences as well as comprehensive knowledge about typical individual, partnership-related, organizational, social and economic aspects and frame conditions will be determined. Content analytic methods will be applied to identify core contents and specific phenomena characterizing the demands, challenges, tasks and scenarios which are important from the view of parents concerning the topic of allergy prevention in early childhood.

### *Developing and pretesting a ECAP-HL item-pool*

Based on the literature review, the expert workshop and the delphi-study as well as the interviews with new parents for each conceptual identified structural component of ECAP-HL a comprehensive item pool will be generated. Items will be formulated as multiple-choice items. Two types of multiple-choice items will be used: *common multiple-choice items* and *scenario based contrastive multiple-choice items*. For *common multiple-choice items* correct response options have to be identified from a number of several preset response options. Two types of multiple-choice items will be used: *common multiple-choice items* and *scenario based contrastive multiple-choice items*. For *common multiple-choice items* correct response options have to be identified from a number of several preset response options. For *scenario based contrastive multiple-choice (SBCM)* items realistic scenarios are formulated, reflecting typical real-life situations. Respondents are requested to read and evaluate both texts. Subsequently, several statements concerning specified literacy aspects are presented. The answer to each statement or the answer pattern on several statements can be coded as “correct” or “incorrect”, because it is defined explicitly which responses are indicative for appropriate literacy. *SBCM item* formats have already been systematically developed and evaluated for the construct Educational Research Literacy (ERL) (18, 34, 35). Such scenario based items proved to exhibit the best content validity, because item contents reflect information or decision situations, which are representative for typical everyday life.

Furthermore, *Situational Judgement Tests (SJT)* (36)) constitute a valid survey format for recording critical prevention-related action patterns. Respondents are asked to judge which of the given behaviors they would prefer depending on the situational conditions. SJT are considered to be an optimal approach to determine the interaction of competence facets and specific framework conditions or tasks. The analysis of possible ways of operationalizing competence facets can thus be extended to include hypotheses on trait-situation interaction. Since new parents have to make flexible use of knowledge and skills in different prevention-relevant situations, this format is valid to examine the construct with regard to its situation dependency. In a first step, the identified scenarios are presented to N = 10 NP using cognitive interviews and comprehension probing (37). These methods are used to ensure that (i) the formulation of the scenarios can be optimized for the study population, and (ii) optimal distractor responses can be developed based on the identified misconceptions and insufficient HL.

The developed items will be rated and annotated by all HELICAP-project groups concerning (1) content validity, (2) construct validity (assigning the item contents to the assumed literacy facets) and (3) comprehensibility. Additionally, cognitive interviews with N = 10 new parents

(think aloud technique, comprehension probing (37)) will be conducted to ensure comprehensibility and content validity of the items.

#### *Analyzing the psychometric properties of the ECAP-HL pilot instrument*

The generated multifaceted item pool will be answered by a sample  $N = 600$  new parents. New parents will be recruited in pediatric practices and with the support of organisations of midwifery care. Study participants will be recruited at all HELICAP-sites and new parents taking part in other HELICAP-projects will also be asked to participate in the WP 8 assessment study

Because of the multifaceted structure of ECAP-HL, the aspired scale reliability and the necessity of deleting non-fitting items from the item pool a large number of about 150 ECAP-HL items is expected. But answering such a high number of test items would overstrain study participants. Accordingly, a block design will be used to minimize strain for participants (38, 39). For this purpose, the itempool will be divided in 6 blocks (A, B, C, D, E, F) with 25 items each. Items of each block are selected equally from all literacy facets and difficulty levels. From these 6 blocks 6 partially overlapping testlets (66% overlap in successive testlets) will be composed. Testlets are defined as follows: testlet 1: A, B, C; testlet 2: B, C, D; testlet 3: C, D, E; testlet 4: D, E, F; testlet 5: E, F, A; testlet 6: F, A, B). Half of every testlet will be organized in block inverse order to minimize position effects (38, 39). Each participant answers one randomly assigned testlet. Despite incomplete data within the generated data set, this design allows for an unbiased psychometric analysis of the whole item pool, because random distribution of testlets ensures a missing-at-random missing data (MCAR) process (40). Comprehensive IRT-modeling is possible, because the partial overlapping testlets ensure a sufficient information linkage within the whole data set (41).

Two samples of at least  $N = 600$  new parents and  $N = 600$  childless adults will take part in the study. Because every participant answers 75 (number of items per testlet) of 150 items, each single item is answered by  $N = 300$  participants in a total sample of  $N = 600$  in each sample. 300 responses per item allow for a stable IRT-solution (30) and sufficiently precise parameter estimates (99% Confidence interval for  $\pm 0,5$  logit units on the according latent trait; (42)).

The obtained raw data will be entered in a SPSS data file. 10% of the data will be entered double to check for data entry errors. A SPSS-syntax device will be adopted to apply coding rules calculating correct vs. incorrect responses from the raw data.

#### *Psychometric modeling and scale analysis of the ECAP-HL pilot instrument*

The psychometric modeling and scale analysis will be conducted in subsequent analytic steps. The theory based multifaceted model structure and item-facet/item-factor mapping will

be defined as confirmatory factor model in the software MPlus 8.2 (43). As data are categorical (correct vs. incorrect) the WLSMV-algorithm (weighted least square mean variances) for dichotomous and ordinal categorical data will be applied. Each item will be defined as a unique indicator of its underlying latent trait. Latent traits are allowed to be intercorrelated. The probability of a correct response will be assumed to correspond to (1) the 1-PL-model (Rasch-model) or (2) the 2-PL-model (Birnbaum-model; IRT-models for dichotomous data, (30)), respectively. The 1-PL-model assumes the item-discrimination to be constant over all items, whereas the 2-PL allows varying item-discriminations. The 1-PL-model has several advantageous properties (e. g. the number of correctly solved items is a sufficient statistic of persons ability). The Bayesian Information Criterion (BIC) and the Consistent Akaike Information Criterion (CAIC) allow determining the appropriate parametric model alternative (Comparative model evaluation).

The RMSEA (root mean square error of approximation) and the SRMR (standardized root mean residual) are measures of the amount of statistical information in the data not explained by the assumed model. In IRT-modeling RMSEA/SRMR-values  $< .1$  indicate an appropriate model fit (44). Item-facet-associations will be evaluated by absolute factor loadings and the significance of factor loadings. If model-fit is not acceptable, misfitting items will be removed from the item pool. If this strategy does not lead to a satisfactory solution, the matrix of unexplained/residual correlations will be inspected and an exploratory factor analysis will be performed in order to identify causes of insufficient model fit.

After consolidation of homogeneous item-groups detailed scale analysis will be performed by the software R and Conquest. The following criteria have to be met (45).

- Weighted and unweighted MNSQ-values in the range [0.8; 1.2]; no significant deviation from the expected value 1.
- Absolute value of Item-Fit residuals below 2.5.
- Item-person-interaction statistics: mean value not significantly deviating from 0, standard deviation not significantly deviating from 1 (46).
- No differential item functioning for potential moderation variables (e. g. kind of allergy, allergy of parents, age and gender of parents or of the child, reading comprehension) (47).
- Less than 5% of significant differences in Smith's t-test (48) comparing items with positive and negative loadings on the first component to prove unidimensionality.
- In an exploratory factor analysis the second main component should explain less than 5% of the overall data variance.

The precision of person estimates will be estimated by the EAP/PV (expected a posteriori/ plausible value) reliability coefficient which represents the explained variance in the estimated model divided by total person variance, and is comparable with Cronbach's  $\alpha$  (43).

Reliability coefficients of .75 or higher are considered as good, although values of at least .55 are deemed satisfactory for group comparisons (49).

The content validity of extracted item groups/assessment scales will be judged by members of all projects in the HELICAP-network. If important aspects concerning content validity are not appropriately represented due to the statistical item selection additional items will be generated and psychometrically tested in the validation assessment.

#### *Moderating and associated constructs of the ECAP-HL pilot instrument*

To analyze the assessments' characteristics for different sub-groups and its association to potential moderating and associated constructs the following sociodemographic and health related characteristic will be answered by the participants:

- Sociodemographic data and health status of the child: age of participant and of the child, gender of the participant and the child, participants marital status, health problems of the child, number of children.
- Socioeconomic status: DEGS-assessment „socioeconomic state“ (50).
- Allergy status and allergy-specific experience of the participant: The “Score for allergic rhinitis (SFAR)-assessment” will be adopted for a broader spectrum of allergies; (51)
- General HL: TOFHLA-assessment (52)
- Reading comprehension: Because items require to read and understand text information the “[reading comprehension test for adults] (LEVE)” (53)
- Allergy-specific experience of threat.
- Allergy-specific knowledge: Higher Knowledge is associated with higher ECAP-HL.
- Allergy-specific Control beliefs: Higher control beliefs are associated with higher ECAP-HL (54).
- Allergy-specific self-efficacy: Higher self-efficacy is associated with higher ECAP-HL (55)
- Allergy-specific prevention motivation
- Allergy-specific risk competence

#### *Validation of the ECAP-HL-Assessment*

The final ECAP assessment will be answered by a sample of N = 300 new parents and N = 300 childless adults. All items will be answered by all parents (complete test design) to determine the psychometric properties of the instrument as recommended for the assessment practice. New parents will be recruited like in the first assessment phase in pediatric practices and with the support of organisations of midwifery care.

To analyze the stability over time of the assessment scales participating new parents will answer the ECAP-HL assessment in two points in time (distance in time 8 weeks). At the first

measurement point in time the same instruments as in the first assessment study will be used (age of participant and of the child, gender of the participant and the child, participants marital status, health problems of the child, DEGS-assessment „socioeconomic status“, adapted version of the “score for allergic rhinitis” (SFAR)-assessment; (51), S-TOFHLA-assessment (52), and the reading comprehension test for adults LEVE (53). At the second measurement point in time general health related attitudes and locus of control will be assessed (54), because these characteristics are theoretically important validation aspects for the ECAP-HL construct: new parents who feel confident that the individual behavior affects health systematically should exhibit specific attitudes and certain behaviors concerning prevention of allergy (55).

### *Psychometric modeling and validation*

The structure of the ECAP assessment and each assessment scale will be analyzed in the same way as in the first assessment phase. Stability over time will be determined by intra class correlation coefficients (56). Bi- and multivariate manifest and latent relationship between theoretically relevant constructs are analyzed using IRT-based statistical evaluation procedures (57, 58) and with a multilevel framework (59, 60).

Validation will focus on the unidimensionality of scales, items-construct associations and the separability of scales. Varying item characteristics depending on potential moderation variables (e. g. kind of allergy, allergy of parents, age and gender of parents or of the child, reading comprehension, locus of control) and over time will be analyzed by means of Differential-Item-Functioning (45, 47).

### *Ethics statement*

The Ethics statement for the original HELICAP *WP Measurement* was approved on January 10<sup>th</sup> 2019 (German Society of Psychology; file number: MAW 112018).

## **1 Bibliography**

1. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H, HLS-EU Consortium. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health*, 2012; 12:80. doi: 10.1186/1471-2458-12-80
2. von Wagner C, Steptoe A, Wolf MS, Wardle J. Health literacy and health actions: a review and a framework from health psychology. *Health Educ Behav*, 2009; 36(5):860-77. doi: 10.1177/1090198108322819
3. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. *American Journal of Health Behavior*. 2007; 31:S19–S26. doi: 10.5555/ajhb.2007.31.supp.S19
4. Parker R. Health literacy: a challenge for American patients and their health care providers. *Health Promot Int*, 2000; 15(4):277-83. doi: 10.1093/heapro/15.4.277

5. Peerson A, Saunders M. Health literacy revisited: what do we mean and why does it matter?. *Health Promot Int*, 2009; 24(3):285-96. doi: 10.1093/heapro/dap014
6. Brown NJS, Furtak EM, Timms M, Nagashima SO, Wilson M. The evidence-based reasoning framework: assessing scientific reasoning. *Educ Assess*, 2010; 15(3/4):123-41. doi: 10.1080/10627197.2010.530551
7. Schladitz S, Groß Ophoff J, Wirtz MA. [Construct validation of a test measuring educational research literacy]. *ZfPaed*, 2015; 61:167-84
8. Westlake C, Sethares K, Davidson P. How can health literacy influence outcomes in heart failure patients? Mechanisms and interventions. *Curr Heart Fail Rep*, 2013; 10(3):232-43. doi: 10.1007/s11897-013-0147-7
9. Pedaste M, Mäeots M, Siiman LA, de Jong T, van Riesen SAN, Kamp ET, Manoli DD, Zacharia TC, Tsourlidadi E. Phases of inquiry-based learning: definitions and the inquiry cycle. *Educ Res Rev*, 2015; 14:47-61. doi: 10.1016/j.edurev.2015.02.003
10. Sørensen K, Pelikan JM, Rothlin F, Ganahl K, Slonska Z, Doyle G, Fullam J, Kondilis B, Agraftotis D, Uiters E, Falcon M, Mensing M, Tchamov K, van den Broucke S, Brand H, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *Eur J Public Health*, 2015; 25(6):1053-8. Doi: 10.1093/eurpub/ckv043
11. Soellner R, Rudinger G. [Health literacy]. In MA Wirtz, C-W Kohlmann, C Salewski (Eds): [Psychology in health promotion], 2017; Hogrefe: Bern, Switzerland:73-84
12. Van den Broucke S. Health literacy: a critical concept for public health. *Arch Public Health*, 2014; 72(1):10. Doi: 10.1186/2049-3258-72-10
13. Nutbeam D. The evolving concept of health literacy. *Soc Sci Med*, 2008; 67(12):2072-8. doi: 10.1016/j.socscimed.2008.09.050
14. Marmot MG, Bell R. Action on health hisparities in the United States. *JAMA*, 2009; 301(11):1169-71. doi: 10.1001/jama.2009.363
15. Manganello JA. Health literacy and adolescents: a framework and agenda for future research. *Health Educ Res*, 2008; 23(5):840-7. doi: 10.1093/her/cym069
16. Mancuso JM. Health literacy: a concept/dimensional analysis. *Nurs Health Sci*, 2008; 10:248-55. doi: 10.1111/j.1442-2018.2008.00394.x
17. Nutbeam D. Health literacy as a public goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int*, 2000, 15(3):259-67. doi: 10.1093/heapro/15.3.259
18. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. *Health Promot Int*, 2005; 20(2):195-203. doi: 10.1093/heapro/dah609
19. Manganello JA. Health literacy and adolescents: a framework and agenda for future research. *Health Educ Res*, 2008; 23(5):840-7. doi: 10.1093/her/cym069
20. Weinert FE. Concept of competence: A conceptual clarification, in defining and selecting key competencies. 2001, Hogrefe & Huber Publishers: Ashland, US:45-65
21. Jordan JE, Buchbinder R, Briggs AM, Elsworth GR, Busija L, Batterham R, Osborne RH. The health literacy management scale (HeLMS): a measure of an individual's capacity to seek, understand and use health information within the healthcare setting. *Patient Educ Couns*, 2013; 91(2):228-35. doi: 10.1016/j.pec.2013.01.013
22. Sørensen K, van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, Kondilis B, Stoffels V, Osborne RH, Brand H, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC Public Health*, 2013; 13:948. doi: 10.1186/1471-2458-13-948
23. Davis TC, Long SW, Jackson RH, Mayeaux EJ, George RB, Murphy PW, Crouch MA. Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Fam Med*, 1993; 25(6):391-5
24. Dumenci L, Matsuyama RK, Kuhn I, Perera RA, Siminoff LA. On the validity of the Rapid Estimate of Adult Literacy in Medicine (REALM) Scale as a measure of health literacy. *Commun Methods Meas*, 2013; 7(2):134-43. doi: 10.1080/19312458.2013.789839

25. Rawson KA, Gunstad J, Hughes J, Spitznagel MB, Potter V, Waechter D, Rosneck J. The METER: A brief, self-administered measure of health literacy. *J Gen Intern Med*, 2009; 25(1):67-71. doi: 10.1007/s11606-009-1158-7
26. Zhang XH, Thumboo J, Fong KY, LI SC. Development and validation of a functional health literacy test. *Patient*, 2009; 2(3):169-78. doi: 10.2165/11314850-000000000-00000
27. Collins SA, Currie LM, Bakken S, Vawdrey DK, Stone PW. Health literacy screening instruments for eHealth applications: a systematic review. *J Biomed Inform*, 2012; 45(3):598-607. doi: 10.1016/j.jbi.2012.04.001
28. McCormack L, Bann C, Squiers L, Berkman ND, Squire C, Schillinger D, Ohene-Frempong J, Hibbard J. Measuring health literacy: A pilot study of a new skill-based instrument. *J Health Commun*, 2010; 15(S2):51-71. doi: 10.1080/10810730.2010.499987
29. Hartig J, Höhler J. Multidimensional IRT-models for the assessment of competences. *Stud Educ Eval*, 2009; 35:57-63. doi: 10.1016/j.stueduc.2009.10.002
30. Van der Linden WJ. (Ed). *Handbook of Item Response Theory, Vol. 3: Applications*, 2018; Chapman and Hall:London, GB
31. Nguyen TH, Han H-R, Kim MT, Chan HS. An introduction to Item Response Theory for patient-reported outcome measurement. *Patient*, 2014; 7(1):23-35. doi: 10.1007/s40271-013-0041-0
32. Groß Ophoff J, Wolf R, Schladitz S, Wirtz MA. Assessment of educational research literacy in higher education: construct validation of the factorial structure of an assessment instrument comparing different treatments of omitted responses. *Journal for Educational Research Online*, 2017; 9(2):35-66
33. Flores-Mateo G, Argimon JM. Evidence based practice in postgraduate healthcare education: a systematic review. *BMC Health Serv Res*, 7:119. doi: 10.1186/1472-6963-7-119
34. Pfaff H, Ernstmann N, Groß SE, Gloede t, Ansmann I, Nitzsche A, Jung J, Wirtz MA, Baumann W, Osburg S, Neumann M. The stressed oncologist in the clinical encounter - results of the WIN ON-study. *Onkol*, 2013; 36:4
35. Müller E, Bengel J, Wirtz MA (2014). [Validity and responsiveness of the Scales for the Assessment of Functioning at Work]. *Rehabilitation (Stuttg)*, 2014; 53(3):176-83. doi: 10.1055/s-0033-1349142
36. Wirtz MA, Strohmer J. [Application and integration of qualitative and quantitative research methods in intervention studies in rehabilitation research]. *Rehabilitation (Stuttg)*, 2016; 55:191-9. doi:10.1055/s-0042-105940
37. Pohontsch N, Meyer T. [Cognitive interviewing – a tool to develop and validate questionnaires]. *Rehabilitation (Stuttg)*, 2015; 54(01):53-59. doi: 10.1055/s-0034-1394443
38. Frey A, Hartig J, Rupp AA. An NCME instructional module on booklet designs in large-scale assessments of student achievement: Theory and practice. *Educ Meas*, 2009; 28(3):39–53
39. Gonzalez E, Rutkowski L. Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. In M von Davier, E. Hastedt (Eds): *IERI monograph series: Issues and methodologies in large-scale assessments*. 2010; Hamburg, Germany, IEA-ETS Research Institute, 3:125–156
40. Schafer JL, Graham JW. Missing data: our view of the state of the art. *Psychol Methods*. 2002; 7(2):147-77. doi: 10.1037/1082-989X.7.2.147
41. Kuhn JT, Kiefer T. Optimal test assembly in practice. The design of the Austrian educational standards assessment in mathematics. *ZfPsych*, 2013; 221(3):190–200. doi: 10.1027/2151-2604/a000146
42. Linacre JM. Sample size and item calibration stability. *Rasch Meas Trans*, 1994; 7:328–31
43. Muthén L, Muthén B. *MPlus: statistical analysis with latent variables: user's guide*, 2010; Muthén&Muthén, Los Angeles, USA



44. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociol Meth Res*, 1992; 21(2):230-58. doi: 10.1177/0049124192021002005
45. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*, 2007; 57, 1358-62. doi: 10.1002/art.23108
46. Covic T, Pallant JF, Conaghan PG, Tennant A. A longitudinal evaluation of the Center for Epidemiologic Studies-Depression scale (CES-D) in a rheumatoid arthritis population using Rasch analysis. *Health Qual Life Outcomes*, 2007;5:41. doi: 10.1186/1477-7525-5-41
47. Zumbo BD. Three generations of differential item functioning (DIF) analyses: Considering where it has been, where it is now, and where it is going. *Lang Assess Q*, 2007; 4, 223–3. doi: 10.1080/15434300701375832
48. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J of Appl Meas*, 2002; 3:205–31
49. Waugh RF. Creating a scale to measure motivation to achieve academically: linking attitudes and behaviours using Rasch measurement. *Br J Educ Psychol*, 2002; 72:65–86. doi: 10.1348/000709902158775
50. Lampert T, Kroll LE, von der Lippe E, Müters S, Stolzenberg H. [Socioeconomic status and health: results of the German Health Interview and Examination Survey for Adults (DEGS1)]. *Bundesgesundheitsbl* 2013; 56:814-21. doi: 10.1007/s00103-013-1695-4
51. Annesi-Maesano I, Didier A, Klossek M, Chanal I, Moreau D, Bousquet J. The score for allergic rhinitis (SFAR): a simple and valid assessment method in population studies. *Allergy*, 2002; 57(2):107-14. doi: 10.1034/j.1398-9995.2002.1o3170.x
52. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J Gen Intern Med*, 10(10):537-41. doi: 0.1007/BF02640361
53. Proyer R, Wagner-Menghin M, Grafinger G. [Test of reading comprehension for adults. – LEVE]. *Schuhfried, Mödling, Austria*; 2010
54. Lohaus A, Schmitt GM. [Assessment of locus of control regarding illness and health]. *Hogrefe, Göttingen, Germany*; 1989
55. Schäfer T, Riehle A, Wichmann HE, Ring J. Alternative medicine and allergies: life satisfaction, health locus of control and quality of life. *J Psychosom Res*, 2003; 55(6):543-6. doi: 10.1034/j.1398-9995.2002.23469.x
56. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*, 2016; 15(2):155-63. doi: 10.1016/j.jcm.2016.02.012
57. Anselmi P, Vidotto G, Bettinardi O, Bertolotti G. Measurement of change in health status with Rasch models. *Health Qual Life Outcomes*, 2015; 13:16. doi: 10.1186/s12955-014-0197-x
58. Rijmen F, de Boeck P, van der Maas HLJ. An IRT model with a parameter-driven process for change. *Pschometika*, 2005; 70(4):16. doi: 10.1007/s11336-002-1047-z
59. Heck RH, Thomas SL. An introduction to multilevel modeling techniques. *Routledge, New York, USA*; 2015
60. Wang J, Luo S. Multidimensional latent trait linear mixed model: an application in clinical studies with multivariate longitudinal outcomes. *Stat Med*, 2017; 36(20):3244-56. doi: 10.1002/sim.7347