Sirkka et al. BMC Nursing (2026) 25:368
https://doi.org/10.1186/s12912-026-04622-y

BMC Nursing

RESEARCH Open Access

Development and psychometric testing of the Competence in Work Ability Risk Management and Analysis (Comp-WARMaA) instrument: a methodological study

Johanna Sirkka[1*], Riitta Suhonen[1,2], Juha Liira[3], Jouko Katajisto[4] and Minna Stolt[1,5]

Abstract

Background Work ability risk management and analysis (WARMA) requires professional competence to ensure effective, consistent, and high-quality occupational health nursing. Occupational health nurses (OHNs) play a central role in identifying and addressing work ability risks; however, no comprehensive instrument currently exists to assess their competence. This study aimed to develop and psychometrically evaluate such an instrument.

Methods A descriptive, exploratory, and methodological design was applied. Instrument development followed a three-phase process comprising theory-based item generation, expert panel evaluations with consensus meetings, and empirical testing. Data were collected using two online surveys including the newly developed Competence in Work Ability Risk Management and Analysis (Comp-WARMaA) instrument. Psychometric properties were examined using classical and modern test theory methods, focusing on content validity, internal consistency, and structural validity.

Results The final Comp-WARMaA instrument comprised 40 items covering four competence domains: objective knowledge, perceived knowledge, attitudes and values, and case-based performance. Content validity was confirmed by experts. Internal consistency was acceptable across subscales. Principal component analysis supported the multidimensional structure, whereas confirmatory factor analyses (CFA) and Rasch analyses yielded mixed findings, indicating the need for further refinement and psychometric testing in independent samples.

Conclusions The Comp-WARMaA instrument demonstrated strong content validity and acceptable internal consistency. Structural analyses yielded mixed findings, indicating the need for further refinement and psychometric testing in independent samples. The instrument provides a structured framework for assessing OHNs' competence in WARMA and may inform systematic competence evaluation and professional development in occupational health nursing.

Clinical trial number Not applicable.

*Correspondence:
Johanna Sirkka
johpelz@utu.fi
Full list of author information is available at the end of the article

Keywords Occupational health nursing, Work capacity evaluation, Clinical competence, Surveys and questionnaires, Psychometrics

Background

Occupational health services (OHS) aim to promote health and work ability by identifying factors that threaten employees' capacity to work [1]. As this objective is central to the effectiveness of OHS, reliable instruments are needed to assess professionals' competence in managing and analysing work ability risks.

Work ability is commonly defined as a dynamic balance between an individual's health, functional capacity, competence, and the demands of work within a given work environment [2, 3]. In OHS, work ability is understood not only as an individual attribute but also as an outcome shaped by organizational, psychosocial, and environmental factors [4].

Occupational health nurses (OHNs) play a central role in promoting work ability, preventing work-related health risks, and supporting safe and sustainable work participation across the working life. Their scope of practice includes health promotion, assessment of work ability, prevention of occupational risks, sickness absence management, and return-to-work support [5, 6]. OHNs work in close collaboration with employers, employees, and multiprofessional occupational health teams at both individual and organizational levels. Within this professional scope, OHNs are well positioned to identify early signals of work ability decline, analyze work-related risk factors, and coordinate preventive and corrective actions with workplaces.

Work ability risk management and analysis (WARMA) constitutes a core component of OHNs' professional responsibilities and plays a key role in supporting and maintaining work ability at both individual and organizational levels [4, 7–9]. In this study, WARMA is used to describe a practice-oriented approach that integrates established elements of work ability promotion, work disability prevention, sickness absence management, and return-to-work support.

WARMA is understood as an integrative and ongoing process encompassing proactive management of work ability, early intervention, sickness absence management with follow-up, and return-to-work support. These components correspond to primary, secondary, and tertiary prevention and are systematically implemented within OHS to support continuous and sustainable work careers [8, 9]. Although WARMA is not a formally established international framework, it draws on widely used work disability prevention and return-to-work models [4] and reflects how these elements are operationalized in Finnish OHS in accordance with national guidance on good occupational health practice [10].

In practice, proactive work ability management focuses on prevention and promotion before early signs of decline emerge, while early intervention aims to prevent further deterioration. Sickness absence management and return-to-work support focus on identifying employees in need of support and facilitating safe and timely return to work through appropriate workplace measures [2, 4, 7]. Within OHS, OHNs play a key role in coordinating work ability–related assessments, preventive actions, and follow-up processes in collaboration with other occupational health professionals and workplaces.

In this study, WARMA is understood as a professional process embedded in routine occupational health practice, whereas competence in WARMA refers to OHNs' capability to effectively enact this process. Competence is conceptualized as a multidimensional professional capability integrating knowledge, performance, attitudes and values, and their application in practice, consistent with established nursing competence frameworks [11, 12].

WARMA is implemented at both the individual employee and client organization levels and is inherently embedded in multiprofessional collaboration within OHS. It also involves close cooperation between OHS and workplaces [8, 9]. At the individual level, WARMA focuses on supporting work ability through assessment, early response, sickness absence management, and return-to-work processes, whereas at the organizational level it involves monitoring work ability indicators, analyzing patterns and root causes, and supporting preventive and corrective actions in cooperation with the employer. Although WARMA is not currently an established theoretical framework, the present study contributes to its conceptual clarification by operationalizing WARMA-related competence into measurable dimensions grounded in established work ability and work disability prevention models. Accordingly, competence in WARMA requires the ability to operate across levels and professional boundaries and to integrate knowledge, values, and practical judgment in complex occupational health contexts, including but not limited to task-level procedures.

Assessing the competence of OHNs, including their ability to manage and analyze work ability risks, is essential for ensuring the quality of care, optimizing resource allocation, and identifying educational and professional development needs [12]. Professional competence, which requires education and training, refers to the ability to perform assigned organizational tasks in accordance with established criteria [13, 14]. Competence has been divided into key theoretical dimensions, including knowledge, performance, skills, attitudes, values [1] and experience [15]. Notably, skills are directly associated with practical performance [16, 17].

Several competence assessment tools are used in occupational health nursing. However, they typically focus on specific domains, such as respiratory protection [18], or provide broad self-assessments of professional competence, such as the AAOHN Self-Assessment Tool [19]. Instruments such as the Work Ability Index assess the client's work ability and do not measure OHN competence [20].

To date, no validated instrument has been identified to assess OHNs' competence in systematically identifying, analyzing, and managing work ability risks across individual and organizational levels, as operationalized in WARMA. Although WARMA can be viewed as a specific competence within the health care service system, it is continuously integrated into routine occupational health practice and therefore constitutes an essential component of general professional competence. Despite this central role, research in this area remains fragmented, and no comprehensive instrument currently exists to measure this competence. An instrument for measuring OHNs' competence in WARMA could provide a structured approach to evaluating and developing competence, thereby promoting consistency and high-quality OHS. Competence assessment can be initiated through self-assessment; however, the incorporation of objective measures broadens the evaluation. Such impartial measures can include knowledge tests for measuring theoretical understanding and case-based assessments for evaluating practical skills and performance. An instrument measuring OHNs' competence in WARMA could also support education and professional development by helping to identify strengths and areas for improvement. In this respect, it may offer valuable input for curricula development, as systematic training in WARMA could further strengthen the professional preparedness of future OHNs. Moreover, the instrument could contribute to the scientific understanding of occupational health nursing by providing a framework for assessing competence in this evolving field.

Despite the central role of WARMA in occupational health nursing, there is currently no validated instrument available to assess OHNs' competence in this area. This gap underpins the rationale for developing the Competence in Work Ability Risk Management and Analysis (Comp-WARMaA) instrument. The aim of this study was to (1) describe the development process of the Competence in Work Ability Risk Management and Analysis (Comp-WARMaA) instrument and to (2) test its validity and reliability.

Methods

Design

The study employed a descriptive, exploratory, and methodological design, adhering to the steps of instrument development (Fig. 1) [21]. The study followed the STROBE guidelines when applicable [22] and used the COSMIN checklist to guide reporting of measurement properties [23].

Development of Comp-WARMaA instrument

The development of the Comp-WARMaA instrument followed an iterative, theory-informed three-phase process conducted in Finland between 2018 and 2025 (Fig. 1). Phases 1 and 2 focused on conceptual clarification and content development through literature review, qualitative inquiry, expert evaluation, and preliminary quantitative analyses to support item refinement. During these phases, WARMA competence was first explored among occupational health professionals (OHP), leading to the construction of Version 1, and subsequently refined and expanded specifically for OHNs to include additional competence dimensions beyond knowledge and skills.

Phase 3 operationalized the theoretically derived structure into a measurable format and evaluated it empirically using confirmatory factor analysis and Rasch analysis. These analyses were conducted to test and refine the proposed competence framework and were not intended to redefine the construct through exploratory modeling. Content validation was conducted in two expert panel rounds at different time points during the overall development period (2018 and 2025). For reporting clarity, the results are presented according to the three-phase development framework. The instrument was developed specifically for the present study and has not been published elsewhere.

Phase 1: Construction of the Comp-WARMaA instrument (2018 – 2022)

In the first phase, development of the Comp-WARMaA instrument was driven by the goal of creating a self-assessment tool for OHPs to evaluate their subjective competence domains (knowledge and skills) in managing and analysing work ability risks [8].

Through deductive reasoning, the theoretical framework for the development of the instrument was derived from relevant literature [e.g., 24, 25] and key legislation governing OHS, including the Occupational Health Care Act [26] and the related Government Decree [27]. National guidance on good occupational health care practice, consistent with international frameworks for occupational health and work ability promotion, also informed the framework [10, 26, 28]. Additionally, the structure of the instrument was informed by prevention

Phase 1: Construction of the Comp-WARMaA Instrument (2018–2022)	Purpose: To operationalize the concept of competence in WARMA in the context of OHC, including OHCPs
	Theoretical framework, conceptual and operational definitions Instrument Version 1: Self-assessed competence domains: knowledge and skills, total 48 items
	Content validity: Expert analysis (n = 10) Content Validity Index, Scale Validity Average Focus group interview (n = 10) Pilot test (n = 11)
	Instrument pre-testing (Data set 1) via electronic survey (n = 169 OHCPs, response rate 10%)
	Construct validity: Principal component analysis, Promax oblique rotation
	Distributions Internal consistency reliability*:* Cronbach's alpha and item analysis
	Conclusion: The psychometric properties of the Comp-WARMaA instrument appear promising

↓

Phase 2: Instrument Refinement and Competence Dimension Expansion (2023–2024)	Purpose: To focus the instrument specifically for OHNs, to expand and refine its content to comprehensively cover two more domains of competence, and to construct new items measuring competence domains and operationalize these domains in a measurable form in WARMA
	Semi-structured individual thematic interview (n = 10 OHNs) Inductive and deductive content analysis to inform the development the content of additional items
	Conclusion: This phase resulted in the identification of previously missing competence domains, which were operationalized into measurable items

↓

Phase 3: Instrument Testing and Competence Assessment (2025)	Purpose: Operationalization of the WARMA-related competence construct; development and psychometric evaluation of the Comp-WARMA instrument; expansion of the competence framework to include knowledge, performance, attitudes, and values	
	Instrument administration (Data set 2) via electronic survey (n = 989 + 839, total 1828 OHNs, response rate 11%) to assess the validity and reliability: Instrument Version 2: Knowledge test (objective), 12 items; Perceived knowledge (self-assessed), 12 items; Attitudes & Values, 12 items; Skills & Performance, 4-part client case: total 40 items)	
	<u>Content validity:</u> Expert analysis (n = 7) Content Validity Index Expert consensus meeting (n = 7)	<u>Construct validity:</u> Pilot test (n = 12) An online survey (n = 989 + 839, total 1828 OHNs, response rate 11%) Confirmatory Factor Analysis Rasch analysis
	Distributions Internal consistency reliability: Cronbach's alpha and item analysis, Kuder Richardson formula (KR-20)	
	Conclusion: The Comp-WARMaA demonstrated promising psychometric properties	

Fig. 1 The Comp-WARMaA instrument development process

categories for promoting work ability, including primary, secondary, and tertiary levels [4, 7].

The development process of the instrument commenced with the identification of its dimensions, followed by item generation, selection of an appropriate response scale, and compilation and critical evaluation of the item pool [29]. The instrument was designed as a norm-referenced measure, enabling respondents' scores to be interpreted relative to those of other respondents in the target group [30]. Competence was divided into two dimensions: (a) knowledge and (b) skills in managing and analysing work ability risks. The content was then structured into four sub-constructs, comprising a total of 48 items: (1) work ability management (15 items), (2) early response (9 items), (3) sickness absence monitoring (11 items), and (4) support for return to work (13 items). Each subscale assessed respondents' (a) knowledge and (b) skills.

The instrument's items were organized by subscales to ensure a balanced and clear structure, with a logical progression that followed the process of managing and analysing work ability risks [31]. Respondents indicated their level of perceived knowledge for each item ($n = 48$) using a four-point Likert scale (1 = poorly to 4 = well) [21, 30]. A higher total score for the summative variables indicated that the respondent (a) had greater knowledge of work ability risk management and analysis and (b) had greater skills in implementing work ability risk management and analysis.

Phase 2: Instrument refinement and competence dimension expansion (2023 – 2024)

In the second phase of the study, the instrument was streamlined to focus on the professional competencies of OHNs in WARMA. The instrument was expanded to include competence areas related to performance, attitudes, and values. Furthermore, the perceived knowledge approach was complemented with an objective assessment of knowledge [17, 31].

Thematic semi-structured interviews with OHNs ($n = 10$) on their experiences and attitudes towards WARMA were conducted using purposive sampling to formulate items assessing perceived knowledge, performance, and attitudes and values. Data analysis was conducted using inductive and deductive content analysis [9].

Instrument Version 2 resulted in a refined set of 40 items covering key aspects of work ability risk management and analysis, addressing both organizational- and individual-level WARMA practices: 12 items assessing objective, norm-based knowledge in WARMA, 12 items assessing subjective, perceived knowledge in WARMA, 12 attitude and value items related to overall WARMA, and a four-part individual client case focusing on case-based performance in WARMA (Table 1). The first part of the instrument, the objective knowledge dimension, and the second part, the perceived knowledge dimension, were structured so that the items in each dimension formed counterpart pairs; items assessing objective knowledge corresponded to counterparts reflecting perceived knowledge or practical knowledge. The components of the instrument reflected the sub-categories of the WARMA process, which included the levels of work ability promotion: the proactive management of work ability, early response, sickness absence management, and return-to-work support. These prevention-oriented levels informed the conceptual organization of the item pool but were not used as analytical categories in the subsequent psychometric analyses.

The full item wording of the Comp-WARMaA instrument is provided in Supplementary File 1 in Finnish, as originally developed, to support transparency and reproducibility. The instrument has not yet been translated or culturally adapted for international use. For the objective knowledge test, item stems and response options are presented without the answer key to preserve item integrity.

The conceptual structure, item distribution across subscales, response formats, and scoring procedures are described in the manuscript (Table 2). For the objective knowledge section, correct responses were coded as 1 and incorrect or "I do not know" responses as 0. For the perceived knowledge and attitudes and values sections, higher visual analogue scale (VAS) scores indicated higher self-assessed competence or stronger endorsement of WARMA-related orientations, and negatively worded items were reverse-coded prior to analysis. VAS scores were treated as continuous variables in the analyses; categorization into low, moderate, and high levels was applied only for descriptive interpretation.

For the case-based performance section, responses were scored based on their rank-order alignment with

Table 1 Conceptual structure of the Comp-WARMaA instrument

Section and competence dimension	Subscale	Work ability promotion level*
Objective knowledge (12 items)	Proactive management of work ability (3 items) Early response (3 items) Sickness absence monitoring (3 items) Return-to-work support (3 items)	Primary–tertiary
Perceived knowledge (12 items)	Proactive management of work ability (3 items) Early response (3 items) Sickness absence monitoring (3 items) Return-to-work support (3 items)	Primary–tertiary
Attitudes and values (12 items)	Norms (3 items) Significance (3 items) Roles (3 items) Ethics (3 items)	Cross-cutting
Case-based performance (four-part case scenario)	Proactive management of work ability Early response Sickness absence monitoring Return-to-work support	Primary–tertiary

*Work ability promotion levels refer to preventive stages within the WARMA framework, including work ability promotion and primary, secondary, and tertiary prevention

Table 2 Structure and response formats of the Comp-WARMaA instrument

Section (total items)	Subscale / component (items)	Response format
Objective knowledge (*n* = 12)	Proactive management of work ability (*n* = 3)	True / False / I don't know
	Early response (*n* = 3)	True / False / I don't know
	Sickness absence monitoring (*n* = 3)	True / False / I don't know
	Return-to-work support (*n* = 3)	True / False / I don't know
Perceived knowledge (*n* = 12)	Proactive management of work ability (*n* = 3)	VAS (1–10)
	Early response (*n* = 3)	VAS (1–10)
	Sickness absence monitoring (*n* = 3)	VAS (1–10)
	Return-to-work support (*n* = 3)	VAS (1–10)
Attitudes and values (*n* = 12)	Norms (*n* = 3)	VAS (1–10)
	Significance (*n* = 3)	VAS (1–10)
	Roles (*n* = 3)	VAS (1–10)
	Ethics (*n* = 3)	VAS (1–10)
Case-based performance (*n* = 4)	Individual client case (four-part case study)	Ranking task (1–10)

The table presents the structure and response formats of the Comp-WARMaA instrument. The full item wording is provided in Supplementary File 1

expert consensus, with greater alignment indicating stronger case-based performance. The instrument includes multiple response formats: (1) objective knowledge assessed using a true/false/I do not know format, (2) perceived knowledge assessed using a VAS (1 = strongly disagree to 10 = strongly agree), (3) attitudes and values assessed using a VAS with the same scale, and (4) case-based performance assessed using a four-part case scenario in which respondents rank multiple response options (a–j) in order (1–10) according to their hypothetical actions. Because the response format of the case-based section differs from the other components, its results are reported separately.

The VAS was used only in the self-report sections, whereas expert panel ratings for content validity were conducted using a 4-point Likert scale in accordance with CVI methodology. This differentiation of item formats aligns with established principles of competence measurement and scale development, where knowledge, orientations, and applied performance represent related but distinct manifestations of professional competence [21].

Phase 3: Comp-WARMaA instrument testing and competence assessment (2025)

Content validation

Content validation was conducted in two expert panel rounds at different time points (2018 and 2025) [32, 33]. For reporting clarity, both rounds are described under Phase 3 in accordance with the three-phase development framework. The instrument structure, derived in earlier phases from literature, legislation, and established work ability and work disability prevention models, was evaluated with emphasis on item clarity, relevance, and applicability. The purpose was to assess and refine item-level content while maintaining the underlying conceptual structure. The expert panels comprised professionals with expertise in WARMA, OHS, and multidisciplinary collaboration in occupational health contexts [34]. The first panel consisted of ten professionals: OHNs (*n* = 2), occupational health physicians (*n* = 3), an occupational health psychologist (*n* = 1), an occupational physiotherapist (*n* = 1), a training specialist (*n* = 1), a researcher (*n* = 1), and an occupational health account manager (*n* = 1). The second panel comprised seven experts: OHNs (*n* = 3), an occupational wellbeing manager (*n* = 1), an occupational safety manager (*n* = 1), a researcher (*n* = 1), and an occupational safety representative (*n* = 1).

Panel members assessed the instrument's content in terms of clarity, relevance, importance of each subscale, and the appropriateness and potential redundancy of items using a 4-point Likert scale (1 = not relevant, 4 = very relevant) [35, 36]. The expert panels met on multiple occasions, during which item wording, terminology, and conceptual clarity were discussed iteratively until shared agreement was reached [37–39].

As a result of the panel evaluations and consensus discussions, minor revisions were made to improve clarity and comprehensibility. These included refinements to background variables, clarification of several items, and modification of one section of the instrument. No items were removed from the instrument.

Pilot testing

The Comp-WARMaA instrument was pilot tested twice. The first pilot was conducted in 2018 with 11 OHPs including OHNs, physicians, psychologists, and physiotherapists from a single OHS unit [4]. The second pilot was conducted in 2025 with 12 OHNs from another OHS unit. The pilot tests aimed to assess the clarity, usability, and response process of the instrument in real-world data collection settings [40, 35]. Specifically, the pilots evaluated the clarity of instructions and response options, the overall functionality of the instrument, and the time required to complete it.

Samples

Participants were recruited from three independent sources during the development and testing phases of the Comp-WARMaA instrument. The first dataset was collected from one OHS provider, while the second was gathered through the Finnish Occupational Health

Nurses' Association mailing list and another OHS provider. The target population comprised Finnish OHNs working in public or private OHS.

Eligible participants were those currently employed as OHNs in Finland and involved in work ability promotion, sickness absence management, and return-to-work processes in OHS, with sufficient Finnish language proficiency to complete the questionnaire. Eligibility was defined to reflect the core practice areas operationalized in the instrument (work ability promotion, sickness absence management, and return-to-work support). In the first dataset, physicians and other OHPs were also invited to participate to support the initial empirical testing phase. For the second dataset, participation was limited to OHNs to evaluate the psychometric properties of the instrument in the intended end-user group. Only professionals currently practicing in OHS were eligible to participate. Students and professionals working outside OHS were not included in the sampling frame. Two datasets were collected to evaluate the internal consistency and structural validity of the Comp-WARMaA instrument. The first dataset was collected in 2018. A total of 1,703 OHPs employed by a single OHS provider were invited through total population sampling [4]. Of these, 169 responded to the survey, and all responses were included in the analyses.

The second dataset was collected in 2025. A total of 1,828 OHNs from the public and private sectors were invited through total population sampling via the Finnish Occupational Health Nurses' Association ($N = 989$) and an OHS provider ($N = 839$). Of these, 200 responded to the survey, and all responses were included in the analyses. Both datasets were gathered through electronic surveys administered using Webropol version 3.0.

Data analysis

Both classical and modern test theory methodological approaches were applied. Statistical data analysis was conducted using Statistical Package for the Social Sciences (SPSS) version 26.0 (IBM SPSS Statistics for Windows, Version 26.0; IBM Corp., Armonk, NY, USA). Datasets 1 and 2 were analyzed using descriptive statistics, including frequencies, percentages, means, ranges, and standard deviations. Each set was screened for completeness prior to analysis. The proportion of missing item responses was low for most items (< 5%), although some showed slightly higher rates. Missing data appeared to be randomly distributed across items, with no systematic patterns observed. No imputation was performed; analyses were based on available data, and the overall impact of missingness on results was considered negligible [41]. Possible violations of model assumptions (e.g. normality, item distribution) were considered and did not meaningfully affect the results [42].

The content validity of the instrument was assessed by expert panels. The item content validity index (I-CVI) was calculated by summing the ratings of 3 and 4 and dividing by the total number of raters. The overall content validity index of the instrument was also assessed using the Scale-Level Content Validity Index based on the average method (S-CVI/Ave). The S-CVI/Ave was calculated by dividing the sum of the item-level content validity index (I-CVI) values by the total number of items [36, 43].

Internal consistency reliability was analyzed using McDonald's Omega, Kuder–Richardson Formula 20 (KR-20) for the dichotomous (correct/incorrect answer) knowledge test, Cronbach's alphas, and item analysis.

Principal component analysis (PCA) with Promax rotation was used to assess the construct validity of Version 1 of the instrument, and the subscales measuring perceived knowledge as well as attitudes and values were examined using Version 2 of the instrument. Exploratory PCA in Version 1 was conducted using data from OHPs to support item refinement, whereas PCA in Version 2 was conducted in the OHN sample. In Version 2, PCA was complemented by confirmatory factor analysis (CFA) and Rasch modeling to evaluate the proposed competence structure. Prior to conducting PCA, the adequacy of the sample and the suitability of the correlation matrix were assessed using the Kaiser–Meyer–Olkin (KMO) measure and Bartlett's test of sphericity. A KMO value ≥ 0.6 is considered acceptable for PCA. Bartlett's test examines whether the correlation matrix significantly differs from an identity matrix, and a p-value < 0.05 indicates that the correlations are sufficiently large for PCA [44].

CFA was performed using Mplus Version 7.11 to test the hypothesized one-factor structure of the subscale measuring objective knowledge. In this model, four observed indicators (Proactive, Early Response, Sickness Absence Management, and Return-to-work) were specified to load onto a single latent factor labelled Knowledge. Each indicator represented the mean score of the three items belonging to the respective knowledge domain. Model fit was evaluated using χ^2, CFI, TLI, RMSEA, and SRMR. CFI and TLI values ≥ 0.95 were interpreted as indicating good fit, whereas values ≥ 0.90 were considered acceptable. RMSEA values ≤ 0.05 were considered indicative of close fit and values ≤ 0.08 as acceptable. SRMR values ≤ 0.08 were regarded as acceptable, in line with commonly applied structural equation modeling criteria [45]. Detailed fit criteria are provided in Supplementary File 2.

Rasch analysis was performed using WINSTEPS version 5.9.1.0 on the 12-item knowledge test of the Comp-WARMaA instrument. Rasch analysis provides psychometric evidence on item and person performance within the measurement model framework [46, 47]. Item and person fit were evaluated using infit mean square (MnSq) statistics, with values between 0.6 and 1.4 considered acceptable [48]. Unidimensionality was examined using principal component analysis of residuals; ≥50% of variance explained by the Rasch measures and a first contrast eigenvalue < 2.0 were interpreted as supporting unidimensionality [46, 47]. Person separation indices ≥ 2.0 were interpreted as indicating adequate discrimination between respondent ability levels [49]. Predefined statistical criteria were used to guide decisions regarding item retention or potential removal during psychometric evaluation. Items not meeting predefined thresholds were evaluated in relation to both statistical performance and conceptual relevance before any modification decisions. Detailed criteria are provided in Supplementary File 2.

The analyses were conducted on two independent datasets collected for the development and validation phases of the instrument. Although the datasets differed in size, both met the methodological requirements for the planned analyses. Previous simulation studies have demonstrated that satisfactory factor recovery can be achieved with moderate sample sizes when essential psychometric assumptions are fulfilled [50, 51]. As the study aimed to develop and psychometrically test an instrument, no formal a priori power analysis was conducted. The sample sizes were considered appropriate in relation to the study aims and methodological recommendations for early-stage instrument development [52, 53, 21]. In addition, sampling adequacy for factor analytic procedures was supported by KMO values and significant Bartlett's tests, which are commonly used criteria to justify the application of PCA and CFA in instrument development research [52, 53].

Ethical considerations

The study adhered to the European Union's General Data Protection Regulation [54] and followed European research ethics practices [55]. Ethical approval was obtained in two phases corresponding to the different stages of the study. Approval for the initial development phase, including conceptual work, expert collaboration, pilot testing, and early data collection, was granted by a university ethics committee in 2017 (approval No. 67/2017). A separate ethical approval was obtained for the later empirical testing phase involving survey data collection from OHNs; this approval was granted by a university ethics review committee in 2022 (approval No. 22/2022). In addition, permissions were obtained from a national professional association and a participating OHS provider.

Participation in the study was voluntary. Informed consent was obtained from all participants, who were provided with information about the study aims, confidentiality, data protection, and their right to withdraw at any time without consequences. Ethical integrity was maintained throughout the research process through adherence to ethical standards, oversight by ethics review committees, and transparent data handling procedures. Participants' autonomy, dignity, and privacy were respected, and the researcher remained available to address questions or provide further clarification when needed [56].

Table 3 Demographic data on respondents

Variable	*n*	Dataset 1	*n*	Dataset 2
Age, years (mean (SD))	169	46.5 (11.0)	199	45.0 (10.3)
Gender, male/female %	169	14 / 86	198	1 / 99
Working experience in OHC, years (mean (SD))	169	12.8 (10.0)	200	13.4 (9.2)
Level of professional education, %				
College	169	17	196	92
Bachelor's degree		44		8
Master's degree		34		0
Other		5		0
Doctoral degree		0		0

Age and working experience are presented as mean (SD). Percentages total 100% within each dataset

Results

Respondents

Dataset 1 yielded 169/1,703 responses (response rate 10%) from OHPs, whereas dataset 2 yielded 200/1,828 responses (response rate 11%) from OHNs. All responses were included in the analyses (Table 3).

Content validity

Content validity was assessed based on evaluations from two expert panels conducted prior to pilot testing. In both panels, the items demonstrated relevance and clarity in relation to the construct, with item-level content validity index (I-CVI) values ≥ 0.78, which were considered acceptable according to established criteria [36, 38]. The predefined retention criterion was I-CVI ≥ 0.78. As all items met this threshold, no items were removed during the expert panel phase. In addition, Scale-Level Content Validity Index values (S-CVI/Ave) were calculated for both expert panel rounds, resulting in 0.96 and 0.92, respectively. These results indicated excellent overall content validity in both rounds. Values ≥ 0.90 are generally considered acceptable for newly developed instruments [42]. Following the panel reviews, consensus meetings were held, during which two background variables were revised for improved clarity, two new background items were added, and one section of the instrument was modified for better comprehensibility. Based on the second panel's consensus meeting, five additional items were clarified; however, no items were removed from the instrument. A subsequent expert consensus meeting was convened to finalize agreement on the revised items, resulting in further modification of four items without any item removals.

Based on the first pilot study, no modifications to the instrument were made before data collection (dataset 1). Based on the second pilot study, one background variable was refined for clarity. No changes were made to the instrument before the main data collection (dataset 2).

Internal consistency

Internal consistency of the knowledge test (KR-20) was 0.724. Additionally, Cronbach's alpha coefficients were acceptable for most subscales, with the subscale measuring perceived knowledge showing α = 0.74. In contrast, the subscale for attitudes and values demonstrated lower reliability (α = 0.56) [21] (Table 4). On the subscale level, item-to-total correlations were calculated, resulting in 0.49–0.88 for dataset 1 and 0.18–0.72 for dataset 2, respectively. One item (36, measuring context-specificity in WARMA, $r = 0.182$) did not meet the recommended item-to-total correlation threshold of $r \geq 0.2$ [57]. However, this item was retained in the instrument due to its content being considered important to the overall construct measured by the instrument. Because the Comp-WARMaA instrument comprises conceptually distinct competence domains measured using different response formats, reliability estimates were calculated and reported at the subscale level instead of a combined total score.

Table 4 Internal consistency of the Comp-WARMaA instrument

Property	Dataset 1 (*n* = 169)	Dataset 2 (*n* = 200)
KR-20 (Knowledge test)	–	0.724
McDonald's Omega (Knowledge test)	–	0.851
Cronbach's alpha (subscales)	0.79–0.94	0.56 – 0.74
Item-to-total correlations (range)	0.49–0.88	–
– Knowledge test	–	0.252–0.564
– Perceived knowledge	–	0.449–0.721
– Attitudes and values	–	0.183–0.562

Table 5 Summary of PCA results

Property	Dataset 1	Dataset 2
Sample size (N)	169	200
Response rate (%)	10	11
KMO	0.94	–
– Perceived knowledge	–	0.880
– Attitudes and values	–	0.714
Bartlett's test of sphericity	$p < 0.001$	$p < 0.001$
Total variance explained	71.7%	56.1–70.6%
Communalities (range)	0.533–0.864	0.306–0.799

KMO = Kaiser-Meyer Olkin measure of sampling adequacy

Construct validity

Principal component analysis

To evaluate the structural validity of the Version 2 Comp-WARMaA instrument, separate PCA were conducted for the subscales measuring perceived knowledge and attitudes and values (Table 5). Sampling adequacy was confirmed by KMO values of 0.880 and 0.714, and Bartlett's tests of sphericity were significant ($p < 0.001$), indicating that the data were suitable for factor analysis [44]. Based on the eigenvalue criterion and Promax oblique rotation, a two-component solution was extracted for perceived knowledge, representing knowledge at both the individual level and organizational level. For attitudes and values, a five-component solution was identified, with each component reflecting a distinct aspect of professional attitudes and values. The component structures accounted for 56.1% and 70.6% of the total variance, respectively. Communalities ranged from 0.306 to 0.799, exceeding the recommended threshold of 0.30 [44]. The results support the multidimensional structure of the instrument and provide evidence for its theoretical foundations.

In an earlier phase of the study, PCA was also conducted on Version 1 of the instrument (Table 5). That analysis yielded seven principal components explaining 71.7% of the total variance, with strong sampling adequacy (KMO = 0.94) and acceptable communalities (0.533–0.864). The identified components aligned with the theoretical domains of the instrument, including early response, proactive risk management, absence monitoring, return-to-work support, and professional values. These findings provided the basis for refining the instrument into its current version.

Confirmatory factor analysis for the knowledge test

The hypothesized model for the knowledge test (Instrument Version 2) was assessed using CFA. The model specified a single latent factor (Knowledge) indicated by four observed variables: Proactive, Early Response, Sickness Absence Management, and Return-to-Work. For model identification, the loading of the first indicator (Proactive) was fixed to 1, and the remaining factor loadings were freely estimated based on the sample data. Figure 2 illustrates the hypothesized CFA model for the knowledge test.

Error of Approximation (RMSEA), and other standard goodness-of-fit indices (Table 6).

The chi-square statistic was significant ($\chi^2(2) = 17.8$, $p < 0.001$), indicating misfit (Table 6). Fit indices showed mixed results: CFI was 0.955 and SRMR 0.036, both within acceptable ranges, whereas RMSEA was 0.198 and TLI 0.864, failing to meet recommended thresholds. Overall, the indices indicated limitations in model fit according to the predefined fit criteria described in

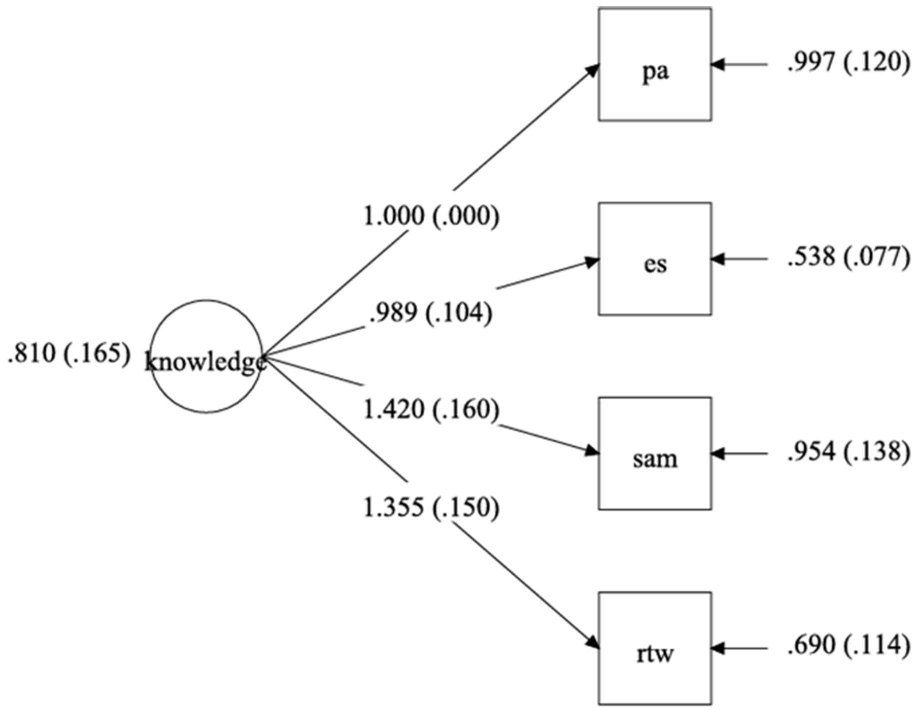

Fig. 2 CFA basic model. p.a. = Proactive, es = Early response, sam = Sickness absence management, rtw = Return-to-work. Model fit was evaluated using the chi-square test, the Comparative Fit Index (CFI), the Root Mean Square

Table 6 Testing (CFA) of the hypothesized structure of the Comp-WARMaA instrument, $n = 200$

Data	χ2	p	d.f.	CFI	TLI	AIC	BIC	SRMR	RMSEA	$p \leq 0.05$
	17.8	< 0.001	2	0.955	0.864	2480	2519	0.036	0.198	< 0.001

CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean Square Error of Approximation; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion

the Methods and based on established structural equation modeling guidelines [45, 58]. In particular, RMSEA exceeded recommended thresholds and TLI remained below acceptable levels, indicating suboptimal fit of the hypothesized model.

The internal consistency of the modeled scale was acceptable (McDonald's Omega = 0.851; Table 4). No indicators were removed. Retention of all four indicators was based on their theoretical relevance to the WARMA knowledge construct.

Rasch analysis for the knowledge test

The Rasch analysis of the knowledge test (Table 7) indicated a mismatch between item difficulty and respondent ability. The mean person measure (3.41 logits) exceeded the mean item difficulty (0.00 logits), suggesting that the items were relatively easy for the sample. Item difficulties ranged from – 2.14 to 3.66 logits.

The total raw variance explained by the Rasch measures was 39.2%, below the commonly referenced 50% threshold for unidimensionality. However, the first contrast eigenvalue was 1.86 (below the 2.0 criterion) and essential unidimensionality reached 61.9%, meeting the ≥ 60% benchmark, in accordance with established Rasch guidelines [46, 47].

Infit MnSq values ranged from 0.74 to 1.29, indicating acceptable item fit, and item reliability was high (0.94). In contrast, person separation was low (real = 0.24; model = 0.57), and person reliability was weak (real = 0.05; model = 0.25), indicating limited discrimination between respondents with higher competence levels.

Marginal DIF was observed for Item 2 by age and work experience; however, no systematic or practically meaningful DIF pattern was identified. No items were removed. Item retention was based on predefined infit, unidimensionality, and reliability criteria together with evaluation of conceptual coverage.

Discussion

Discussion of the results

This study produced the Comp-WARMaA instrument to assess OHNs' competence in WARMA through a systematic, multi-phase development and testing process using both classical and modern psychometric methods.

Table 7 Summary of Rasch analysis results of the knowledge test

Domain	Metric	Value	Range/Group	Interpretation	Notes
Item hierarchy	Mean item difficulty	0.00 logits	-	Reference mean	
	Mean person ability	3.41 logits	-	Items too easy	Consider adding harder items
	Item difficulty range	-2.14 to 3.66	Items 1–10	Broad range	
Rating scale	Category 0 (n, %)	274 (11%)	Low ability	Functioning well	Mean measure = -0.10
	Category 1 (n, %)	2119 (89%)	High ability	Functioning well	Mean measure = 2.79
	MnSq (Infit/Outfit)	1.00 / 0.76	Category 0	Acceptable	
	MnSq (Infit/Outfit)	1.02 / 1.16	Category 1	Acceptable	
Unidimensionality	Rasch variance explained	39.2%	-	Below recommended 50% threshold	Suggests limited proportion of variance explained
	1st contrast eigenvalue	1.86	-	No secondary dimension	Below 2.0 threshold
	Essential unidimensionality	61.9%	-	Meets ≥ 60% criterion	
Reliability	Cronbach's alpha	0.68	-	Moderate consistency	
	Item reliability	0.94	-	Excellent	
	Item separation	4.06	-	Good item spread	
	Person reliability	0.05 (real)	-	Weak	Low person variance
	Person separation	0.24 (real)	-	Low discriminative ability	
DIF	Age-related DIF	Item 2: $p = 0.0439$	Marginal	No systematic bias	
	Work experience DIF	Item 2: $p = 0.0424$	Marginal	No systematic bias	

DIF = Differential Item Functioning; MnSq = Infit mean square

Note: Cronbach's alpha (0.68) differs slightly from the KR-20 estimate (0.724) reported earlier because it was calculated within the Rasch analysis output and may reflect different handling of missing responses

The instrument was developed in line with established approaches to competence instrument development in clinical nursing and occupational health contexts [59–61]. By integrating objective knowledge testing, self-assessed competence, attitudes and values, and performance-based scenarios, the Comp-WARMaA reflects a broad and practice-relevant conceptualization of professional competence in occupational health nursing.

The psychometric findings should be interpreted as an empirical test of a theoretically specified competence framework and not as a data-driven redefinition of the construct. The Comp-WARMaA instrument was developed through an iterative process in which conceptual clarification and content development preceded psychometric evaluation. Accordingly, the competence domains assessed in this study represent complementary dimensions of a single underlying WARMA-related competence, reflecting different manifestations of professional capability and not independent or competing constructs. While some psychometric indicators suggested limitations in model fit, these findings should be interpreted in the context of early-stage instrument development, where iterative refinement and further testing across settings and populations are expected.

Given the importance of WARMA competence in OHS [8, 9] and the preventive focus in the working-age population [2, 4, 62], the instrument provides a means to systematically measure and evaluate a competence area that has received little research attention. Identifying gaps may support targeted professional development and workforce planning in OHS.

Content validity was demonstrated through high I-CVI and S-CVI/Ave scores (> 0.90) in both phases. Multidisciplinary expert panels, consensus meetings, and pilot tests further confirmed item clarity and relevance, aligning with previous recommendations emphasizing expert engagement in instrument development [34, 36]. Internal consistency was acceptable across most subscales: the knowledge test (KR-20 = 0.724), perceived knowledge (α = 0.740), and McDonald's Omega for the knowledge scale (Ω = 0.851) were within recommended limits [57, 21]. The subscale measuring attitudes and values, however, showed lower reliability (α = 0.56). Unlike the other subscales, it was not aligned with the theoretical structure of the other subscales (Proactive, Early response, Sickness absence management, and Return-to-work) but instead targeted overall WARMA orientations, including norms, significance, roles, and ethics. This broader conceptual scope may explain its multidimensionality and weaker internal consistency, a phenomenon also observed in studies examining self-rated clinical competence [63]. The Attitudes and values subscale was designed to capture broader professional orientations related to WARMA, including norms, roles, and ethical considerations. As these elements represent conceptually related but not necessarily homogeneous aspects of professional competence, lower internal consistency was anticipated. While alternative factor structures and item refinement were considered during instrument development, the subscale was retained as a single construct to reflect its conceptual breadth. Further refinement and potential subdivision into more specific dimensions will be explored in future validation studies. Future research could examine whether a multidimensional modeling approach would better capture the attitudinal domain, given its broader conceptual scope.

Construct validity was supported by PCA, whereas CFA and Rasch analyses yielded mixed findings indicating limitations in model fit and measurement precision and requiring further psychometric testing in independent samples. PCA confirmed the multidimensional structure with theoretically coherent components. CFA of the knowledge test yielded mixed fit indices. While the CFI (0.955) and SRMR (0.036) indicated acceptable to good fit, the RMSEA (0.198) substantially exceeded recommended thresholds and the TLI (0.864) remained below the conventional criterion of 0.90, based on commonly applied cutoff criteria in structural equation modeling [45, 58]. The elevated RMSEA therefore indicates clear model misfit under conventional guidelines and warrants cautious interpretation of the one-factor solution. The four observed indicators were conceptualized as reflective manifestations of an underlying knowledge competence domain, consistent with the theoretical specification of WARMA-related competence.

Although RMSEA is known to be sensitive in models with very small degrees of freedom and a limited number of indicators [45], the present findings suggest that the hypothesized structure does not fully capture the covariance structure of the observed variables. Alternative model specifications were examined; however, no theoretically meaningful and statistically superior solution was identified. The one-factor model was retained on theoretical grounds in line with the predefined WARMA competence framework. Given the small number of indicators and the theory-driven specification of the construct, retaining the model was considered preferable to post hoc modifications lacking conceptual justification. Accordingly, the CFA findings should be interpreted as preliminary structural support within a theory-driven framework and not as confirmation of strict unidimensionality.

The Rasch analysis demonstrated high item reliability and acceptable item fit. However, limitations were observed in targeting and person separation. The mean person measure exceeded the mean item difficulty, indicating that the items were relatively easy for the study sample and that discrimination between higher competence levels was limited. Person separation reliability was low, suggesting restricted ability of the instrument to distinguish between respondents with higher competence levels [49, 50].

Although the first contrast eigenvalue remained below 2.0 and essential unidimensionality met the ≥ 60% criterion, the total variance explained by the Rasch dimension (39.2%) was below the commonly referenced 50% threshold for unidimensionality [46, 47]. These findings indicate mixed evidence regarding dimensionality and measurement precision. From a measurement development perspective, the results suggest that refinement of the item pool, including the addition of more demanding items, may improve targeting and discrimination across competence levels [21, 50]. Further validation in independent samples is required to confirm dimensionality, targeting, and person separation. In line with recommendations for scale development, model fit indices should be interpreted in relation to theoretical coherence, construct definition, and the purpose of the instrument, not as absolute decision criteria [21, 57]. The retained one-factor solution was theoretically grounded in the proposed WARMA competence framework, despite overall limitations in model fit, including a high RMSEA and low TLI, although some incremental fit indices (CFI and SRMR) were within acceptable ranges [45, 29].

Mixed fit patterns have been reported in early-stage instrument development, particularly when modeling complex professional competence constructs with a limited number of indicators and a theoretically specified structure [21]. The Rasch analysis provided complementary evidence, demonstrating strong item reliability and acceptable unidimensionality. However, limitations in person separation indicate that further refinement is needed to improve discrimination across competence levels [50, 47].

In addition to its measurement purpose, the Comp-WARMaA instrument contributes to clarifying WARMA as a competence area in occupational health nursing. Although WARMA is not an established theoretical framework, the instrument operationalizes the concept by translating key elements of work ability promotion, work disability prevention, sickness absence management, and return-to-work support into defined competence dimensions. This supports a more explicit understanding of how OHNs enact work ability risk management and analysis in practice at individual and organizational levels.

Overall, the Comp-WARMaA shows promise for use in professional development, recruitment, and continuing education, as well as in evaluating intervention outcomes and informing curricula by identifying competence strengths and gaps. Future research should refine items, test discriminative validity across known groups, and conduct cross-cultural validation to enhance generalizability [64, 65]. Expanding items may ensure coverage of evolving competencies in occupational health nursing [21, 66].

Strengths and limitations

One of the strengths of this study is its systematic, stepwise development process. It combined theory, empirical data, and psychometric analyses across multiple phases. Multidisciplinary expert input, including panel evaluations, consensus meetings, and pilot tests, strengthened content validity. Sample sizes in both datasets ($n = 169$ and $n = 200$) were considered adequate for early-stage psychometric analyses when interpreted in relation to methodological recommendations for instrument development and factor analysis [52, 53, 21]. The samples included participants from both public and private OHS, representing a broad cross-section of Finnish OHNs. With an estimated 2,500–2,600 OHNs working in Finland [67, 68], and with the national association including members from both public and private sectors, the recruitment strategy enabled participation from a broad range of occupational health contexts.

This study has limitations. Response rates were relatively low (10–11%), which may have introduced response bias, as those more interested in WARMA may have been more likely to respond. Anonymity prevented comparison between respondents and non-respondents. Further, the study was conducted in a single national context, which may limit its generalizability, as cultural, organizational, and systemic factors affect competence assessments [21]. The inclusion of physicians and other professionals in dataset 1 may explain the differences between the two datasets, such as dataset 1 having a higher proportion of respondents with master's degrees (34% vs. 8%); this reflects Finland's educational structure where physicians are educated at the master's level and OHNs at the bachelor's level [69, 67, 70]. These differences highlight the importance of context-specific validation and caution in interpreting the findings.

In addition, no a priori power analysis was performed. The sample sizes were considered adequate based on methodological recommendations for exploratory and confirmatory factor analyses and Rasch modeling in instrument development studies, where moderate sample sizes are commonly accepted when key model assumptions are met [51–53]. The analyses focused on examining construct validity and reliability in early-stage instrument development and were not designed for hypothesis testing. The use of two independent datasets and complementary psychometric approaches strengthens the overall evaluation despite the low response rates. A further strength is the use of multiple reliability indices (Cronbach's alpha, KR-20, and McDonald's Omega), which provided a robust evaluation of internal consistency. The subscale measuring attitudes and values, nevertheless, showed lower reliability ($\alpha = 0.56$). Its broad conceptual scope, intended to capture overall orientations such as norms, significance, roles, and ethics, not a single domain, may explain both its multidimensionality and lower consistency. Further item development may strengthen this construct in future versions. Evidence from multiple sources provided partial support for the validity and reliability of the Comp-WARMaA, while also identifying areas requiring further refinement. By systematically identifying competence gaps, the instrument may support professional development, curriculum planning, and workforce management in OHS.

These limitations should be considered when interpreting the findings and the current status of the Comp-WARMaA instrument. Although Version 2 is supported by initial empirical evidence in terms of content validity and overall reliability, some psychometric findings, including limited item difficulty and mixed model fit indices, indicate that further refinement is needed. Version 2 should therefore be regarded as an instrument with initial evidence of validity that remains under development, not as a finalized measure. Additional testing in larger, more diverse, and international samples is required before broader application.

Recommendations for further research

Future research should refine items, test discriminative validity, and evaluate the Comp-WARMaA instrument in larger and more diverse occupational health contexts. In addition, longitudinal studies could explore the relationship between competence levels measured by the instrument and outcomes such as work ability promotion, client satisfaction, and OHS effectiveness. Further studies could also examine the instrument's cross-cultural applicability and measurement invariance in international settings [66, 42]. Exploring how competence in work ability risk management and analysis develops over time and how it relates to nurses' educational background or continuing professional development would also strengthen the evidence base [71]. Moreover, educational intervention studies could be designed to evaluate how targeted training programs based on the Comp-WARMaA framework enhance OHNs' competence and confidence in work ability risk management. Finally, qualitative research could provide a deeper understanding of how OHNs conceptualize competence in work ability management, complementing quantitative findings and supporting ongoing refinement of the instrument [21, 23].

Implications for policy and practice

The instrument offers practical utility for use in professional development, workforce planning, and research within OHS. In addition to its potential for use in research and competence evaluation, the Comp-WARMaA instrument could be applied in practical contexts such as OHN recruitment, continuing education planning, and structured support for work ability management.

In the clinical context of occupational health nursing, the instrument provides a theoretically grounded, standardized framework for assessing and guiding nurses' competence in managing and analysing clients' work ability, sickness absence, and return-to-work processes. By systematically identifying strengths and gaps in WARMA-related competence, its use supports evidence-based competence development and contributes to organizational learning within OHS.

From a broader policy and professional perspective, the Comp-WARMaA enhances the conceptual and empirical basis for evaluating occupational health nursing competence. It promotes comparability and transparency across occupational health settings, thereby strengthening the quality, accountability, and visibility of occupational health nursing as a distinct professional domain. Its systematic use may facilitate benchmarking, inform quality improvement initiatives, and support the harmonization of competence standards across different occupational health systems.

Conclusions

Through the systematic development and testing process of the Comp-WARMaA instrument, this study produced a novel tool designed to assess OHNs' competence in WARMA. The instrument integrates objective knowledge testing, self-assessed competence, attitudes, values, and performance-based scenarios, grounded in both theoretical and practice-based frameworks.

The Comp-WARMaA demonstrated strong content validity and acceptable internal consistency. Structural analyses yielded mixed findings, indicating the need for further refinement and psychometric testing in independent samples. By operationalizing competence in WARMA into measurable dimensions, the instrument provides a structured framework for assessing OHNs' competence in work ability risk management and analysis. With further development, it may inform systematic competence evaluation, professional development, and quality improvement within occupational health nursing.

Abbreviations

CFA	Confirmatory factor analysis
CVI	Content Validity Index
KMO	Kaiser–Meyer–Olkin measure of sampling adequacy
OHN	Occupational health nurse
OHS	Occupational health service(s)
PCA	Principal component analysis
SPSS	Statistical package for the social sciences
VAS	Visual analogue scale
WARMA	Work ability risk management and analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12912-026-04622-y.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

The authors gratefully acknowledge the financial support of the Jenny and Antti Wihuri Foundation. The authors also wish to thank all the experts who contributed to the development of the instrument, all those who participated in the practical implementation of the study, and all respondents who took the time to complete the survey.

Author contributions

JS was the main author of the manuscript and was involved in all aspects of the study. MS, RS, and JL contributed to the design and planning of the study, the data interpretation, and the writing of the manuscript. JK contributed to performing the statistical analyses. MS and RS supervised the study, and all authors (JS, RS, JL, JK, MS) read and approved the final manuscript.

Funding

This work was supported by the Jenny and Antti Wihuri Foundation [grant number 00240356].

Data availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. The full item wording of the Comp-WARMaA instrument is provided in Supplementary File 1. The Comp-WARMaA instrument is protected by copyright (© Johanna Sirkka) and is available for research purposes from the corresponding author.

Declarations

Ethics approval and consent to participate

The study received ethical approvals from the Ethics Committee for Human Sciences at the University of Turku, Health Care Division, Finland (statement number 67/2017, meeting date 11 December 2017, and statement number 22/2022, meeting date 6 June 2022). All participants provided written informed consent prior to participation. The study was conducted in accordance with the Declaration of Helsinki [72] and the national guidelines for the ethical principles of research with human participants and ethical review in human sciences in Finland [73].

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Nursing Science, University of Turku, Medisiina B, Kiinamyllynkatu 10, Turku 20014 Turun yliopisto, Finland
[2]The Wellbeing Services County of Southwest Finland, Turku University Hospital, Turku, Finland
[3]Department of Clinical Medicine, University of Turku, Turku, Finland
[4]Department of Mathematics and Statistics, University of Turku, Turku, Finland
[5]Wellbeing Services County of Satakunta, Pori, Finland

Received: 6 November 2025 / Accepted: 1 April 2026
Published online: 10 April 2026

References

1. World Health Organization. Occupational Health [Internet]. Geneva: World Health Organization: WHO. 2019. [cited 2025 Aug 15]. Available from: https://www.who.int/health-topics/occupational-health
2. Ilmarinen J, Rantanen J. Promotion of work ability during ageing. Am J Ind Med. 1999;36(Suppl 1):21–3. 10.1002/.(SICI)1097-0274(199909)36:1+<21::AID-AJIM8>3.0.CO;2-S.
3. Ilmarinen J. Work ability—a comprehensive concept for occupational health research and prevention. Scand J Work Environ Health. 2009;35(1):1–5. https://doi.org/10.5271/sjweh.1304.
4. Loisel P, Anema JR, editors. Handbook of work disability: Prevention and management. New York (NY): Springer; 2013.

5. World Health Organization. The role of the occupational health nurse in workplace health management [Internet]. Geneva: World Health Organization. 2001 [cited 2026 Jan 6]. Available from: https://www.who.int/publications/i/item/the-role-of-the-occupational-health-nurse-in-workplace-health-management
6. Federation of Occupational Health Nurses within the European Union (FOHNEU). Definition of occupational health nursing [Internet]. Brussels: FOHNEU; c2024 [cited 2026 Jan 6]. Available from: https://fohneu.org/about-us/ohn-definition
7. Institute for Work & Health. Primary, secondary and tertiary prevention [Internet]. Toronto: Institute for Work & Health; 2015 [cited 2025 Aug 15]. Available from: https://www.iwh.on.ca/what-researchers-mean-by/primary-secondary-and-tertiary-prevention
8. Sirkka J, Suhonen R, Katajisto J, Salminen L, Stolt M. Self-reported competence level of occupational health care professionals in work ability risk management and analysis. Work. 2023;76(3):1071–81. https://doi.org/10.3233/WOR-220229.
9. Sirkka J, Suhonen R, Liira J, Stolt M. Occupational Health Nurses' Perceptions in Work Ability Risk Management and Analysis. J Occup Rehabil. 2025;35:423–33. https://doi.org/10.1007/s10926-025-10282-7.
10. Uitti J, Pokela T, Lemiläinen L-M. Hyvä työterveyshuoltokäytäntö. 3. täysin uudistettu painos. Helsinki: Sosiaali- ja terveysministeriö; 2014. [Finnish].
11. Cowan DT, Norman I, Coopamah VP. Competence in nursing practice: a controversial concept—a focused review of literature. Nurse Educ Today. 2005;25(5):355–62. https://doi.org/10.1016/j.nedt.2005.03.002.
12. Meretoja R, Isoaho H, Leino-Kilpi H. Nurse competence scale: development and psychometric testing. J Adv Nurs. 2004;47(2):124–33. https://doi.org/10.1111/j.1365-2648.2004.03071.x.
13. Le Deist FD, Winterton J. What is competence? Hum Resour Dev Int. 2005;8(1):27–46. https://doi.org/10.1080/1367886042000338227.
14. Zumstein-Shaha M, Grace PJ. Competency frameworks, nursing perspectives, and interdisciplinary collaborations for good patient care: delineating boundaries. Nurs Philos. 2023;24:e12402. https://doi.org/10.1111/nup.12402.
15. Bergen J, Santo L. Action competence: a concept analysis. Nurs Forum. 2018;53:358–63. https://doi.org/10.1111/nuf.12263.
16. Cheetham G, Chivers G. The reflective (and competent) practitioner: a model of professional competence which seeks to harmonize the reflective practitioner and competence-based approaches. J Eur Ind Train. 1998;22(7):267–76.
17. Fukada M. Nursing competency: definition, structure and development. Yonago Acta Med. 2018;61(1):1–7. https://doi.org/10.33160/yam.2018.03.001.
18. Taormina D, Burgel BJ. Development of a respiratory protection survey instrument for occupational health nurses: an educational project. Workplace Health Saf. 2013;61(2):79–83. https://doi.org/10.1177/216507991306100206.
19. Burgel BJ. Self-assessment tool for occupational and environmental health nurses. Am Assoc Occup Health Nurses. 2003.
20. Ilmarinen J. The work ability index (WAI). Occup Med (Lond). 2007;57(2):160. https://doi.org/10.1093/occmed/kqm008.
21. DeVellis RF, Thorpe CT. Scale development: theory and applications. 5th ed. Thousand Oaks (CA): SAGE; 2022.
22. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. Int J Surg. 2014;12(12):1500–24. https://doi.org/10.1016/j.ijsu.2014.07.014.
23. Mokkink LB, Terwee CB, Patrick DL, et al. COSMIN study design checklist for patient-reported outcome measurement instruments [Internet]. Amsterdam: COSMIN; 2019 [cited 2025 May 30]. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf.
24. Taimela S, Justén S, Aronen P, Sintonen H, Läärä E, Malmivaara A. An occupational health intervention programme for workers at high risk for sickness absence: cost effectiveness analysis based on a randomised controlled trial. Occup Environ Med. 2008;65(4):242–8. https://doi.org/10.1136/oem.2007.033167.
25. Schmidt L, Sjöström J, Antonsson A-B. How can occupational health services in Sweden contribute to work ability? Work. 2012;41(Suppl 1):2998–3001. https://doi.org/10.3233/WOR-2012-0555-2998.
26. Finland, Ministry of Social Affairs and Health. Occupational Health Care Act (No. 1383/2001) [Internet]. Unofficial translation. Helsinki: Ministry of Social Affairs and Health; 2001 [cited 2025 May 6]. Available from: https://www.finlex.fi/api/media/statute-foreign-language-translation/688546/mainPdf/main.pdf.
27. Finland. Government Decree on the principles of good occupational health care practice, the content of occupational health care and the educational qualifications required of professionals and experts (No. 708/2013) [Internet]. Helsinki: Ministry of Social Affairs and Health; 2013 [cited 2025 May 6]. Available from: https://www.finlex.fi/fi/lainsaadanto/saadoskaannokset/2013/eng/708.
28. World Health Organization (WHO). Occupational health: a manual for primary health care professionals [Internet]. Geneva: World Health Organization. 2020 [cited 2025 May 6]. Available from: https://www.who.int/publications/i/item/occupational-health---a-manual-for-primary-health-care-workers
29. Noar SM. The role of structural equation modeling in scale development. Struct Equ Model. 2003;10(4):622–47. https://doi.org/10.1207/S15328007SEM1004_8.
30. Schmidt JD, Register-Mihalik JK, Mihalik JP, Kerr ZY, Guskiewicz KM. Identifying impairments after concussion: normative data versus individualized baselines. Med Sci Sports Exerc. 2012;44(8):1621–8. https://doi.org/10.1249/MSS.0b013e318258a9fb.
31. Garside JR, Nhemachena JZZ. A concept analysis of competence and its transition in nursing. Nurse Educ Today. 2013;33(5):541–5. https://doi.org/10.1016/j.nedt.2011.12.007.
32. Schilling LS, Dixon JK, Knafl KA, et al. Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. Nurs Res. 2007;56(5):361–6. https://doi.org/10.1097/01.NNR.0000289505.30037.91.
33. Weston MJ, Verran JA, Clavelle JT, Porter-O'Grady T. Professional governance scale instrument development and content validity testing. Adv Nurs Sci. 2018;41(2):188–98. https://doi.org/10.1097/ANS.0000000000000200.
34. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res Nurs Health. 1997;20(3):269–74. https://doi.org/10.1002/(sici)1098-240x(199706)20:3%26lt;269::aid-nur9%26gt;3.0.co;2-g.
35. Pittman J, Bakas T. Measurement and instrument design. J Wound Ostomy Cont Nurs. 2010;37(6):603–7. https://doi.org/10.1097/WON.0b013e3181f90a60.
36. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health. 2006;29(5):489–97. https://doi.org/10.1002/nur.20147.
37. Rattray J, Jones MC. Essential elements of questionnaire design and development. J Clin Nurs. 2007;16(2):234–43. https://doi.org/10.1111/j.1365-2702.2006.01573.x.
38. Vogt DS, King DW, King LA. Focus groups in psychological assessment: enhancing content validity by consulting members of the target population. Psychol Assess. 2004;16(3):231–43. https://doi.org/10.1037/1040-3590.16.3.231.
39. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and implementation content validity study: development of an instrument for measuring patient-centered communication. J Caring Sci. 2015;4(2):165–78. https://doi.org/10.15171/jcs.2015.017.
40. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol. 2010;10:67. https://doi.org/10.1186/1471-2288-10-67.
41. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken (NJ): Wiley; 2014.
42. Polit DF, Beck CT. Nursing research: generating and assessing evidence for nursing practice. 11th ed. Philadelphia (PA): Wolters Kluwer; 2021.
43. Rodrigues IB, Adachi JD, Beattie KA, MacDermid JC. Development and validation of a new tool to measure the facilitators, barriers and preferences to exercise in people with osteoporosis. BMC Musculoskelet Disord. 2017;18:540. https://doi.org/10.1186/s12891-017-1914-5.
44. Child D. The essentials of factor analysis.3rd ed. London: Continuum International; 2006.
45. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
46. Lerdal A, Kottorp A, Gay CL, Aouizerat BE, Lee KA, Miaskowski C. A Rasch analysis of assessments of morning and evening fatigue in oncology patients using the Lee Fatigue Scale. J Pain Symptom Manage. 2016;51(6):1002–12. https://doi.org/10.1016/j.jpainsymman.2015.12.331.
47. Stolt M, Kottorp A, Suhonen R. The use and quality of reporting of Rasch analysis in nursing research: a methodological scoping review. Int J Nurs Stud. 2022;132:104244. https://doi.org/10.1016/j.ijnurstu.2022.104244.
48. Smith RM, Schumacker RE, Bush MJ. Using item mean squares to evaluate fit to the Rasch model. J Appl Meas. 2003;4(2):205–31.
49. Fisher WP Jr. Reliability, separation, strata statistics [Internet].Rasch Meas Trans.1992;6(3):238 [cited 2025 Aug 23]. Available from: https://www.rasch.org/rmt/rmt63i.htm.

50. Bond TG, Fox CM. Applying the Rasch model: fundamental measurement in the human sciences. 3rd ed. New York (NY): Routledge; 2015.
51. MacCallum RC, Widaman KF, Zhang S, Hong S. Sample size in factor analysis. Psychol Methods. 1999;4(1):84–99. https://doi.org/10.1037/1082-989X.4.1.84.
52. Mundfrom DJ, Shaw DG, Ke TL. Minimum sample size recommendations for conducting factor analyses. Int J Test. 2005;5(2):159–68. https://doi.org/10.1207/s15327574ijt0502_4.
53. Costello AB, Osborne JW. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Pract Assess Res Eval. 2005;10(7):1–9.
54. European Parliament and Council of the European Union. Regulation (EU) 2016/679: general data protection regulation (GDPR) [Internet]. Brussels: EUR-Lex. 2016 [cited 2025 Aug 26]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679
55. ALLEA. The European code of conduct for research integrity. Revised ed. Berlin: ALLEA. 2023 [cited 2025 Aug 26]. Available from: https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf
56. Beauchamp TL, Childress JF. Principles of biomedical ethics. 8th ed. New York (NY): Oxford University Press; 2019.
57. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 3rd ed. Oxford: Oxford University Press; 2003.
58. Weston R, Gore PA, Chan F, Catalano D. An introduction to using structural equation models in rehabilitation psychology. Rehabil Psychol. 2008;53(3):340–56. https://doi.org/10.1037/a0013039.
59. Haavisto E, Siltanen H, Tolvanen A, Holopainen A. Instruments for assessing healthcare professionals' knowledge and skills of evidence-based practice: a scoping review. J Clin Nurs. 2023;32(15–16):4391–407. https://doi.org/10.1111/jocn.16506.
60. Chuang JY, Liang HY, Dai HD, Tseng TY, Yu S. A professional nurse competence scale (PNCS): development and psychometric testing. BMC Nurs. 2025;24:624. https://doi.org/10.1186/s12912-025-03231-5.
61. Billiau L, Taghon D, Duprez V, Eeckloo K, Malfait S. Instruments for measuring head nurses' competencies in a hospital setting: a scoping review. BMC Nurs. 2025;24:1069. https://doi.org/10.1186/s12912-025-03719-0.
62. World Health Organization. Health workforce requirements for universal health coverage and the Sustainable Development Goals (Human Resources for Health Observer, No. 17)[Internet]. Geneva: World Health Organization. 2016 [cited 2025 Sep 30]. Available from: https://www.who.int/publications/i/item/9789241511407
63. Vikström-Dahl I, Bing-Jonsson PC, Rauhala A, Fagerström L. Nursing standards, language and age as variables associated with clinical competence for nurses in long-term care facilities: a cross-sectional study. J Clin Nurs. 2024;33(11):4381–94. https://doi.org/10.1111/jocn.17102.
64. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186–91. https://doi.org/10.1097/00007632-200012150-00014.
65. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health. 2018;6:149. https://doi.org/10.3389/fpubh.2018.00149.
66. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res. 2018;27(5):1159–70. https://doi.org/10.1007/s11136-018-1829-0.
67. International Labour Organization. National profile of occupational health system in Finland [Internet]. Geneva: International Labour Office. 2004 [cited 2025 Oct 23]. Available from: https://www.ilo.org/media/320116/download
68. Nissinen S, Oikarinen T, Lappalainen K, et al. Occupational health services: activities and quality in Finland 2024 [Internet]. Helsinki: Finnish Institute of Occupational Health; 2024 [cited 2025 Oct 23]. Available from: https://www.julkari.fi/handle/10024/149959. [In Finnish].
69. Finland. Health Care Professionals Act (No. 559/1994) [Internet]. Helsinki: Ministry of Social Affairs and Health; 1994 [cited 2025 May 6]. Available from: https://www.finlex.fi/api/media/statute-foreign-language-translation/688923/mainPdf/main.pdf.
70. World Health Organization (WHO). Finland – strengthening health systems through nursing: evidence from 14 European countries (Health Policy Series, No. 52) [Internet]. Copenhagen: WHO Regional Office for Europe. 2019 [cited 2025 May 6]. Available from: https://www.euro.who.int/en/publications/abstracts/finland-strengthening-health-systems-through-nursing-2019.
71. Taylor I, Bing-Jonsson P, Wangensteen S, Finnbakk E, Sandvik L, McCormack B, Fagerström L. The self-assessment of clinical competence and the need for further training: a cross-sectional survey of advanced practice nursing students. J Clin Nurs. 2020;29(3–4):545–55. https://doi.org/10.1111/jocn.15095.
72. World Medical Association. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. JAMA. 2013;310(20):2191–4. https://doi.org/10.1001/jama.2013.281053.
73. Finnish National Board on Research Integrity TENK. The ethical principles of research with human participants and ethical review in the human sciences in Finland. Helsinki: TENK. 2019. [cited 2025 Nov 11]. Available from: https://tenk.fi/en/advice-and-materials.

Publisher's note

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