ClearingHouse
Finanzbildungsforschung

CHF Open Access Repository

Documentation No. 1/2025

German Pension System Understanding Scales (GePSUS)

Ronja Baginski

Citation:

Baginski, R. (2025). *German pension system understanding scales (GePSUS).* CHF Open Access Repository (No. 1/2025). https://www.uni-mannheim.de/media/Einrichtungen/chf/Dokumente/CHF_OAR_2025_1_GePSUS_Baginski.pdf

MetaFin

UNIVERSITY OF MANNHEIM ZEW IÖB INSTITUT FÜR ÖKONOMISCHE BILDUNG

Terms of use:

Content

1 Summary

The GePSUS (German pension system understanding scales) is a test to assess adults' understanding of the German pension system with three scales. The first scale measures the understanding of the statutory pension insurance (SPI scale), the second the understanding of a private pension insurance (PPI scale), and the third the understanding of the factors that influence the stability of the German pension system and its components (Factors scale).
The test has the following characteristics:

Test name:	German pension system understanding scales (GePSUS)
Author:	Ronja Baginski
CHF-OAR No.:	1/2025 (in the repository since November 19, 2025)
Measured construct:	German pension system understanding with three scales Scale 1: Understanding of the statutory pension insurance (SPI scale) Scale 2: Understanding of a private pension insurance (PPI scale) Scale 3: Understanding of factors influencing the stability of the pension system (Factors scale)
Number of items:	21 test items (in sum), each scale has the following number of items: Scale 1 (SPI): 11 items, Scale 2 (PPI): 7 items, Scale 3 (Factors): 3 items
Scoring:	Max. 21 points, Min. 0 points Correct response (underlined) = 1, other responses = 0
Target group:	Adults (aged between 18 and 68 years and living in Germany)
Test use:	Large-scale surveys
Languages:	Instrument: German, Documentation: English
Mode:	CATI (Computer Assisted Telephone Interviews)
Development:	Based on the Evidence-Centered Assessment Design (ECD) according to Mislevy and Riconscente (2005), see Baginski (2025) for details
Validation:	Based on the Standards for Educational and Psychological Testing (AERA et al., 2014), see Baginski (2025) for details Validity: Evidence was collected regarding (1) test content, (2) response processes, (3) internal structure, and (4) relation to other variables using content analysis as well as methods from Classical Test Theory (CTT) and Item-Response Theory (IRT) Reliability: According to IRT modelling (Testinformation), the SPI and PPI scales measure rather low levels and the Factors scale rather medium level of understanding very precisely Fairness: Only 4 items showed DIF when evaluating fairness regarding the variables *Migration background, East/West German state* of residence, *Gender*, *Age*, and *Education level*
Funding:	The development and the validation of the test were a part of the project VHAlt (Verstaendnis und Haltungen zur Altersvorsorge in Deutschland: Auspraegungen und Auswirkungen auf vorsorgebezogenes Verhalten) (Aprea & Übelmesser, 2018), which was funded by the BMAS (Bundesministerium fuer Arbeit und Soziales) as a part of the "Foerdernetzwerk interdisziplinäre Sozialpolitikforschung" (No. FIS.00.00011.19).

2 Instrument

GePSUS (German pension system understanding scales)

General instructions to the respondents:
Im folgenden Abschnitt werde ich Ihnen nun einige Fragen zu unterschiedlichen Formen der Altersvorsorge in Deutschland stellen. Falls Sie eine Antwort auf eine Frage mal nicht wissen, geben Sie dies an. Versuchen Sie aber bitte, bei den kommenden Fragen immer erst zu überlegen.

Instructions to the interviewee:
Do not read out loud the response options “Weiß nicht” and “Keine Angabe”.
Choose “Weiß nicht” if a respondent does not know the answer and “Keine Angabe” if the respondent refuses to answer.

Scale 1: Understanding the statutory pension insurance (SPI)
Instructions to the respondents:
Als erstes geht es um die gesetzliche Rentenversicherung.

Scale 1 (SPI) – Part 1/3:
Instructions to the respondents:
Wir interessieren uns für die Beitragszahlungen in die gesetzliche Rentenversicherung. Bitte geben Sie an, ob die folgenden Aussagen Ihrer Meinung nach zutreffen oder nicht.
[Not randomised]

No.	German wording (English version included in the appendix)	Response options (correct ones underlined)
1.1	In Deutschland sind alle Personen, die einer selbstständigen oder nicht selbstständigen Erwerbstätigkeit nachgehen, zu Beitragszahlungen in die gesetzliche Rentenversicherung verpflichtet.	1 Trifft zu 2 <u>Trifft nicht zu</u> 3 Weiß nicht 999 Keine Angabe
1.2	In Deutschland werden die Beitragszahlungen in die gesetzliche Rentenversicherung in voller Höhe von den Arbeitnehmern getragen.	1 Trifft zu 2 <u>Trifft nicht zu</u> 3 Weiß nicht 999 Keine Angabe
1.3	In Deutschland sind Beamte zu Beitragszahlungen in die gesetzliche Rentenversicherung verpflichtet.	1 Trifft zu 2 <u>Trifft nicht zu</u> 3 Weiß nicht 999 Keine Angabe

Scale 1 (SPI) – Part 2/3:
Instructions to the respondents:
Und was geschieht Ihrer Meinung nach mit den Beiträgen, die in die gesetzliche Rentenversicherung eingezahlt werden? Bitte geben Sie wieder an, ob die folgenden Aussagen Ihrer Ansicht nach zutreffen oder nicht.
[Not randomised]

2.1	Die Beiträge werden für jeden Versicherten von der gesetzlichen Rentenversicherung auf einem gesicherten Konto angespart.	1 Trifft zu 2 <u>Trifft nicht zu</u> 3 Weiß nicht 999 Keine Angabe

2.2	Die Beiträge werden durch die gesetzliche Rentenversicherung am Kapitalmarkt angelegt.	1 2 3 999	Trifft zu Trifft nicht zu Weiß nicht Keine Angabe
2.3	Die Beiträge werden verwendet, um eine gerechtere Verteilung zwischen Rentnern mit geringem Einkommen und solchen mit hohem Einkommen herbeizuführen.	1 2 3 999	Trifft zu Trifft nicht zu Weiß nicht Keine Angabe
2.4	Die Beiträge werden verwendet, um die Rentenzahlungen an die Personen zu finanzieren, die aktuell in Rente sind.	1 2 3 999	Trifft zu Trifft nicht zu Weiß nicht Keine Angabe
Scale 1 (SPI) – Part 3/3: Jetzt geht es um die Rentenhöhe. Bitte sagen Sie mir, ob die folgenden Faktoren Ihrer Ansicht nach einen großen Einfluss darauf haben, wie viel Rente jemand aus der gesetzlichen Rentenversicherung erhält. Bitte antworten Sie mit Ja, wenn Ihrer Ansicht nach ein großer Einfluss vorliegt, und mit Nein, wenn dies nicht der Fall ist. [Randomised]			
3.1	Die Art und Weise, wie die Deutsche Rentenversicherung das Geld anlegt, hat einen großen Einfluss darauf, wie viel Rente jemand aus der gesetzlichen Rentenversicherung erhält.	1 2 3 999	Ja Nein Weiß nicht Keine Angabe
3.2	Die Entwicklung der Zinsen am Kapitalmarkt hat einen großen Einfluss darauf, wie viel Rente jemand aus der gesetzlichen Rentenversicherung erhält.	1 2 3 999	Ja Nein Weiß nicht Keine Angabe
3.3	Die Höhe des letzten Einkommens vor Renteneintritt der versicherten Person hat einen großen Einfluss darauf, wie viel Rente jemand aus der gesetzlichen Rentenversicherung erhält.	1 2 3 999	Ja Nein Weiß nicht Keine Angabe
3.4	Die Verwaltungskosten der Deutschen Rentenversicherung haben einen großen Einfluss darauf, wie viel Rente jemand aus der gesetzlichen Rentenversicherung erhält.	1 2 3 999	Ja Nein Weiß nicht Keine Angabe
Scale 2: Understanding a private pension insurance (PPI)			
Scale 2 (PPI) – Part 1/2 Instructions to the respondents: Nun wenden wir uns den privaten Rentenversicherungen als weiterer Form der Altersvorsorge in Deutschland zu. Bitte geben Sie wieder an, ob die folgenden Aussagen aus Ihrer Sicht zutreffen oder nicht. [Not randomised]			
4.1	Bei einer privaten Rentenversicherung wird ein Geldbetrag bei einer Versicherungsgesellschaft eingezahlt, der zu Rentenbeginn in eine monatliche Rente umgewandelt werden kann.	1 2 3 999	Trifft zu Trifft nicht zu Weiß nicht Keine Angabe

4.2	Eine private Rentenversicherung ist für Selbstständige verpflichtend, wenn diese nicht in die gesetzliche Rentenversicherung einzahlen.	1 Trifft zu 2 Trifft nicht zu 3 Weiß nicht 999 Keine Angabe
4.3	Um aus einer privaten Rentenversicherung eine Rente in geplanter Höhe zu erhalten, gilt: Je länger man einzahlt, desto geringer kann der Betrag sein, den man im Monat einzahlen muss.	1 Trifft zu 2 Trifft nicht zu 3 Weiß nicht 999 Keine Angabe
Scale 2 (PPI) – Part 2/2 Und was hat Ihrer Ansicht nach einen großen Einfluss auf die Höhe der Rente, die jemand aus einer privaten Rentenversicherung erhält? Ich lese Ihnen wieder einige Faktoren vor. Bitte antworten Sie mit Ja oder Nein. [Randomised]		
5.1	Die Art und Weise, wie die jeweilige Versicherungsgesellschaft das Geld anlegt, hat einen großen Einfluss darauf, wie viel Rente jemand aus einer privaten Rentenversicherung erhält.	1 Ja 2 Nein 3 Weiß nicht 999 Keine Angabe
5.2	Die Entwicklung der Zinsen am Kapitalmarkt hat einen großen Einfluss darauf, wie viel Rente jemand aus einer privaten Rentenversicherung erhält.	1 Ja 2 Nein 3 Weiß nicht 999 Keine Angabe
5.3	Die Höhe des letzten Einkommens vor Renteneintritt der versicherten Person hat einen großen Einfluss darauf, wie viel Rente jemand aus einer privaten Rentenversicherung erhält.	1 Ja 2 Nein 3 Weiß nicht 999 Keine Angabe
5.4	Die Verwaltungskosten der jeweiligen Versicherungsgesellschaft haben einen großen Einfluss darauf, wie viel Rente jemand aus einer privaten Rentenversicherung erhält.	1 Ja 2 Nein 3 Weiß nicht 999 Keine Angabe
Scale 3 (Factors) – Part 1/1 Bestimmte langfristige Veränderungen können die Stabilität des gesamten Systems der Altersvorsorge beeinflussen. Ich lese Ihnen nun drei solcher Veränderungen vor. Wenn sich sonst nichts anderes ändert, wie wirkt sich die jeweilige Veränderung auf die Stabilität der gesetzlichen beziehungsweise einer privaten Rentenversicherung aus? [Randomised]		
6.1	Wirkt sich ein sinkendes Zinsniveau langfristig stärker auf die Stabilität der gesetzlichen Rentenversicherung oder die Stabilität einer privaten Rentenversicherung oder auf beide etwa gleich stark oder auf keine von beiden aus?	1 Stärker auf die gesetzliche Rentenversicherung 2 Stärker auf eine private Rentenversicherung 3 Auf beide etwa gleich stark 4 Auf keine von beiden 5 Weiß nicht 999 Keine Angabe
6.2	Wirkt sich ein Anstieg der Arbeitslosenquote langfristig stärker auf die Stabilität der gesetzlichen Rentenversicherung oder die Stabilität einer	1 Stärker auf die gesetzliche Rentenversicherung 2 Stärker auf eine private Rentenversicherung

	privaten Rentenversicherung oder auf beide etwa gleich stark oder auf keine von beiden aus?	3 Auf beide etwa gleich stark 4 Auf keine von beiden 5 Weiß nicht 999 Keine Angabe
6.3	Wirkt sich ein Anstieg der Inflationsrate langfristig stärker auf die Stabilität der gesetzlichen Renten-versicherung oder die Stabilität einer privaten Rentenversicherung oder auf beide etwa gleich stark oder auf keine von beiden aus?	1 Stärker auf die gesetzliche Rentenversicherung 2 Stärker auf eine private Rentenversicherung 3 Auf beide etwa gleich stark 4 Auf keine von beiden 5 Weiß nicht 999 Keine Angabe

3 Theoretical background

The GePSUS was based on a theoretical background that consists of three research areas, which are described in detail in Baginski (2025). The first area is research in systems science (Ackoff, 1994; Meadows, 2008). Of particular importance for the GePSUS was research on the characteristics of complex systems such as the interrelatedness between a system's components and the whole (Ladyman et al., 2013; Estrada, 2023). This was transferred to the German pension system to conceptualise it as a complex system. Furthermore, research on fostering the ability to understand such complex systems, i.e., systems thinking (Evagorou et al., 2009), is another important basis because the GePSUS aims at measuring the understanding of the German pension system. As a part of this research on systems thinking in science education, Hmelo-Silver et al. (2007, 2017) developed the Components-Mechanisms-Phenomenon (CMP) framework to break down systems into interacting components, i.e. mechanisms, and resulting phenomena. This approach was transferred to the German pension system to conceptualise it consisting of components, mechanisms, and phenomena.

The second area refers to research on lay understanding in science and social science as the test aimed at operationalising the lay understanding of adults living in Germany. Thus, research on lay thinking compared to expert thinking was considered (Kahneman, 2011) as well as research on lay understanding particularly of economic phenomena (Aprea, 2015; Davies & Lundholm, 2012; Leiser & Shemesh, 2018). This research came to similar findings regarding the lay view on systems as research in systems science. For example, laypersons do not consider feedback effects (Leiser & Shemesh, 2018) but assume linear relationships (Jacobson & Wilensky, 2006).

The third area is research in educational science on knowledge and understanding. It was considered to define the kind of knowledge that the GePSUS is supposed to measure. Referring to the taxonomy of Anderson et al. (2014), the GePSUS aimed at operationalising conceptual knowledge because making sound long-term decisions in a pension system requires not only knowledge about facts or the systems components (factual knowledge), but also about the relationship between these components, i.e., the mechanisms.

In sum, this theoretical background served as a basis to define the construct the GePSUS aims to measure, i.e., *pension system understanding* defined as understanding the complex structure of a pension system with its subsystems and interacting components as well as the mechanisms and resulting phenomena (Baginski, 2025, p. 46).

4 Development

This section provides a short overview of the development process of the GePSUS and the empirical studies that were conducted. Details are described in Baginski (2025).

The GePSUS was developed following the framework of Evidence-Centered Assessment Design (ECD) (Mislevy & Riconscente, 2005). The ECD consists of five layers: *domain analysis, domain modeling, conceptual assessment framework, assessment implementation*, and *assessment delivery*. During *domain analysis,* information about the domain was collected based on three different empirical studies, i.e., 1) a literature review, 2) the analysis of newspaper comments, and 3) an interview study.

In the first study, a review of specialised literature on the German pension system was conducted in 2019. This served to further operationalise the construct to be measured, i.e., understanding of the German pension system. The specialised literature review aimed at grasping the expert view on the functioning of the German pension system. Thus, books from pension experts (e.g., Koehler-Rama, 2018, 2020), research articles, websites from relevant pension institutions such as the German statutory pension insurance agency (DRV, 2024) as well as legal documents were analysed. This information was then structured according to the CMP framework (Hmelo-Silver et al., 2007, 2017; Hmelo-Silver & Pfeffer, 2004). Based on the proposed three-pillar structure (DRV, 2024), the German pension system was conceptualised consisting of three components, which are the statutory pension insurance, occupational pensions, and private pension options such as a private pension insurance. These components are based on two different functioning schemes, i.e., mechanisms. The first scheme, on which the German statutory pension insurance is based on, is the pay-as-you-go scheme. Here, contributions paid in by working and insured employees are directly paid out as pensions to retirees. This scheme stands in contrast to the funded model underlying private and occupational pensions, where each insured individual finances their own entitlements (Koehler-Rama, 2020). This is important as the GePSUS is supposed to measure not only understanding of the German pension system structure but also understanding of the two functioning schemes. Furthermore, different factors such as population ageing or inflation have an impact on the financial stability of the whole pension system that should also be understood by laypersons so that they are able to make wise financial decisions for the long-term. Thus, based on a review of the specialised literature, the study indicated that the GePSUS should include three components: the three-pillar structure, the two pension schemes, and the main influencing factors.

The second study consisted of an analysis of newspaper comments with the aim of capturing laypersons' views on the German pension system and identifying misunderstandings that the GePSUS should address. Therefore, 240 comments to articles about pensions from four newspapers with differing readership and political orientation were selected, downloaded, and analysed in 2019 (for more details see Baginski, 2025). As a result, the analysis revealed which topics are important for laypersons and where misunderstandings occur, such as the misconception about how pensions are calculated in the statutory pension insurance.

The third study was a semi-structured interview study conducted in 2020. It served several purposes. Firstly, it helped to make further decisions on the covered topics and content in the GePSUS. Secondly, a variety of questions with open-ended responses were developed that build the basis for constructing test items with a closed response format. Thirdly, further misunderstandings were identified to then design a test that operationalises them. Finally, it helped to develop a test that is well understood by laypersons, e.g., regarding the wording. The interview sample consisted of 24 adults living in Germany. They were selected

based on *Profession, Gender, Age, Education,* and a residence in *East/West German state* aiming for maximum heterogeneity. A detailed description of the sample is included in Baginski (2025, p. 54). The interviews were conducted online using Zoom (due to the COVID-19 pandemic), audio-recorded, transcribed verbatim, and anonymised. The transcripts were analysed using MAXQDA 2020 and coded using a coding guideline that was developed based on Mayring (2015) and referring to earlier category systems developed for coding interview material on informal conceptions of economic phenomena (Aprea, 2015) and on pensions (Baginski et al., 2021). As a result, different levels of understanding and several misunderstandings, e.g., of the statutory pension insurance and its pay-as-you-go scheme, of its contribution payers or pension receivers were identified. Further details on the results are described in Baginski (2021). The responses to the open-ended questions were used to develop items with a closed response-format for the GePSUS in the following ECD layers.

The collected information in the ECD layer of *domain analysis* was then systematically organised in the *domain modeling*. Referring to the structure of arguments developed by Toulmin (1958, 2003) and the design patterns of Mislevy and Riconscente (2005), a test concept was developed. This concept includes a description of the test taker, the test user, the testing context, the measured constructs, rationales, as well as the used psychometric theories and the item format (Baginski, 2025, pp. 60-62).

In the third layer, i.e., *conceptual assessment framework*, further decisions regarding the design of the GePSUS were made to develop a test blueprint. Decisions were made concerning the scales, scoring procedures, measurement model, response format, testing context, the number of items required, and the mode of presentation.

In the *assessment implementation*, concrete parts of the GePSUS such as the wording of items and instructions were developed based on the framework from the former ECD layer. Thus, a first test version was available after this layer. The fifth and final layer, i.e., *assessment delivery*, includes administering the GePSUS to test takers to investigate how they respond to it. As this marks the transition to the validation process of the GePSUS, this is described in the next section.

5 Validation

The validation process of the GePSUS was developed in line with the Standards for Educational and Psychological Testing (AERA et al., 2014). Four empirical validation studies were conducted to form an argument of validity, i.e., to collect *validity evidence* based on the *test content,* the *response processes,* the *internal structure,* and the *relation to other variables.* In addition, it was investigated whether the GePSUS meets the criteria of *reliability/precision* and *fairness.* The four empirical validation studies comprised 1) expert feedback, 2) cognitive interviews, 3) a pre-test, and 4) a survey. Details are described in Baginski (2025).

The first validation study was conducted in fall 2020 to gather *validity evidence* based on *test content*. Feedback was collected on the initial version of the test from three experts selected due to their knowledge of the German pension system and of survey design. The experts provided comments on the domain coverage, wording, format, and scoring of the items and instructions, using the comment function in Microsoft Word. The experts' comments were analysed, and the test was revised, respectively. Afterwards, the results were presented to and discussed with the research team of the project VHAlt. Further changes were undertaken, which resulted in the second test version.

The second test version was used in a cognitive interview study in October 2020 to collect *validity evidence* based on the *response processes* by using a qualitative method. The sample consisted of 12 young adults. A description of the sample is included in Baginski (2025, p. 86). The interviews were conducted via Zoom (due to the COVID-19 pandemic). In line with Pruefer and Rexroth (2005), the interviewees were asked to respond to the items thinking aloud. The audio-recorded interviews were transcribed verbatim and anonymised. The content was analysed focussing particularly on potential misunderstandings regarding the formulation of test instructions and items. Afterwards, several changes to the GePSUS were undertaken regarding the response options and instructions as well as the wording. This resulted in a third test version.

The third version of the GePSUS was used in the third validation study, a pre-test conducted in November 2020 by a professional survey company. The study aimed to collect additional *validity evidence* based on *response processes* using quantitative methods. It was decided to use the mode of computer assisted telephone interviews (CATI) for the pre-test and the following survey (validation study 4) due to the following reasons: As the respondents are asked the test items per phone, they do not have the option to look up correct responses simultaneously as easily as in online surveys. Furthermore, the rate of drop out during the test as well as skipping items was supposed to be lower and the interviewees were able to ask questions to the interviewer. The pre-test was conducted with 30 adults living in Germany. A detailed description of the sample is included in Baginski (2025, p. 89). The data of the pre-test was analysed regarding item difficulty according to CTT (see Baginski, 2025, p. 91-93). Items that were too easy or too difficult were dropped from the scale, which resulted in a fourth test version of the GePSUS.

The fourth test version was then used in a survey with 1,000 adults living in Germany. A sub-sample of N = 200 was selected for validation purposes. Both samples are described in Baginski (2025, p. 224, p. 101). The data of the validation sample was then analysed to collect the remaining *validity evidence* and to check whether the GePSUS is a *reliable/precise* and *fair* assessment instrument. Details on this are provided in Baginski (2025, Section 7.4). In short, five analysis steps were conducted, which combine methodological approaches from Classical Test Theory (CTT) and Item-Response Theory (IRT), as recommended in the related literature

(Bean & Bowen, 2021; Moosbrugger et al., 2020). The first step aimed at collecting *validity evidence* based on the *response processes* and the *internal structure.* Thus, item analyses were conducted, in which the item difficulties and the discrimination indexes were calculated according to CTT. Furthermore, IRT models (Rasch and 2PL) were calculated to elicit the difficulty parameter and the discrimination parameter. Afterwards, correlations and Cronbach's α were calculated to gain first insights into the *internal structure* of the GePSUS. Based on these results, decisions were made regarding the drop of items.

Afterwards, a Confirmatory Factor Analysis (CFA) was conducted to analyse the internal structure of the GePSUS. The results of this analysis basically confirmed the three factors, i.e., the SPI understanding (Scale 1), the PPI understanding (Scale 2), and the understanding of the factors (Scale 3), with a few exceptions, that are explained in more detail in Baginski (2025).

In a third step, the three scales were Rasch modelled (Rasch, 1960). Based on the item characteristic curves (ICCs) and the test information, it was shown that the first two scales, i.e., understanding the statutory pension insurance and understanding a private pension insurance, measure rather low levels of understanding very precisely while the third scale, i.e., understanding the factors influencing the stability of the pension system, measures medium level of understanding very precisely. Moreover, it was found that all three scales can be Rasch modelled so that sum scores can be calculated and used in the following analysis step.

In the fourth step, correlations of the three sum scores were calculated to collect further *validity evidence* based on the *internal structure* of the GePSUS as well as to collect *validity evidence* based on the *relation to other variables*. The sum scores of the three scales correlated significantly positive with each other, which provided *validity evidence* based on the *internal structure*. Furthermore, correlations between the three scale scores and the variables *Interest, Education level,* and *Self-confidence* were investigated to collect *convergent evidence*. This evidence was provided partly by significantly positive correlations regarding the SPI scale and the PPI scale and the variables *Education level* and *Self-confidence*. *Interest* did not correlate significantly with any of the three scale scores. In addition, the PPI scale score did not correlate with any of the three variables. Thus, *validity evidence* based on the *relation to other variables* was provided to a certain extent, but further investigations are needed regarding the PPI scale score as well as regarding *divergent evidence*.

In the fifth and final analysis step, the fairness of the GePSUS and its scales regarding the sociodemographic variables *Gender, Age, Education level, Migration background,* and *East/West German state* was assessed using Differential Item Functioning (DIF) (Wetzel & Boehnke, 2017). For this, the *R* package *DifR* developed by Magis et al. (2010) was used to conduct five different tests for DIF simultaneously. Only four items showed DIF (Item 3.4 regarding *Gender*, Item 1.3 regarding *Age* and *Education level,* and Item 5.3 regarding *Migration background*). Therefore, further investigations on the reasons for the DIF as well as reformulations of the respective items need to be undertaken in the future. Until then, when analysing differences between the respective sociodemographic groups, the respective item should be excluded from the scale, which reduces the total test score by 1 point.

In conclusion, the validation process of the GePSUS provided sufficient *validity evidence* as well as support for *reliability/precision* and *fairness*. The instrument effectively assesses pension system understanding across its three scales: understanding of the statutory pension insurance (Scale 1: SPI), understanding of a private pension insurance (Scale 2: PPI), and understanding the factors influencing the stability of the pension system (Scale 3: Factors).

6 References

Ackoff, R. L. (1994). Systems thinking and thinking systems. *System Dynamics Review*, *10*(2–3), 175–188.

AERA (American Educational Research Association), APA (American Psychological Association), & NCME (National Council on Measurement in Education). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Anderson, L. W., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., & Wittrock, M. (2014). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's*. Pearson.

Aprea, C. (2015). Secondary school students' informal conceptions of complex economic phenomena. *International Journal of Educational Research*, *69*, 12–22. https://doi.org/10.1016/j.ijer.2014.09.002

Aprea, C., & Uebelmesser, S. (2018). *Verstaendnis und Haltungen zur Altersvorsorge in Deutschland: Auspraegungen und Auswirkungen auf vorsorgebezogenes Verhalten (VHAlt)* [Understanding and attitudes towards old age provision in Germany: Characteristics of and consequences for precautionary behaviour] [Unpublished project proposal]. University of Mannheim.

Baginski, R. (2021, December 2). *Lay conceptions of old age provision in Germany: an interview study on understanding and attitudes* [Paper presentation of an unpublished working paper]. Early Career Workshop of the Mannheim Institute for Financial Education (MIFE), Mannheim (online), Germany.

Baginski, R. (2025). *Development and validation of a test measuring pension system understanding in Germany* [Dissertation]. University of Mannheim.

Baginski, R., Schultheis, J., & Aprea, C. (2021, August 23). *Young adults' intuitive conceptions of the pension system in Germany: An interview study* [Presentation]. 19th Biennal European Association for Research on Learning and Instruction (EARLI) Conference, Gothenburg (online), Sweden.

Bean, G. J., & Bowen, N. K. (2021). Item response theory and confirmatory factor analysis: Complementary approaches for scale development. *Journal of Evidence-Based Social Work*, *18*(6), 597–618. https://doi.org/10.1080/26408066.2021.1906813

Davies, P., & Lundholm, C. (2012). Students' understanding of socio-economic phenomena: Conceptions about the free provision of goods and services. *Journal of Economic Psychology*, *33*(1), 79–89. https://doi.org/10.1016/j.joep.2011.08.003

DRV (Deutsche Rentenversicherung Bund). (2024). *Die drei Saeulen der Altersvorsorge* [The three pillars of old-age provision]. Retrieved March 25, 2024, from https://www.deutsche-rentenversicherung.de/DRV/DE/Rente/Moeglichkeiten-der-Al-tersvorsorge/Drei-Saeulen-der-AV/Drei-Saulen-der-AV_node.html.

Estrada, E. (2023). What is a complex system, after all? *Foundations of Science*, *29*, 1143–1170. https://doi.org/10.1007/s10699-023-09917-w

Evagorou, M., Korfiatis, K., Nicolaou, C., & Constantinou, C. (2009). An investigation of the potential of interactive simulations for developing system thinking skills in elementary school: A case study with fifth-graders and sixth-graders. *International Journal of Science Education*, *31*(5), 655–674.

Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: A quasi-experimental study. *Instructional Science*, *45*, 53–72. https://doi.org/10.1007/s11251-016-9392-y

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *The Journal of the Learning Sciences*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*(1), 127–138. https://doi.org/10.1016/S0364-0213(03)00065-X

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the Learning Sciences*, *15*(1), 11–34.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Koehler-Rama, T. (2018). *Das Rentensystem verstehen: Einfuehrung in die Politische Oekonomie der Alterssicherung* [Understanding the pension system: Introduction into the political economy of old-age provision]. Wochenschau Verlag.

Koehler-Rama, T. (2020). *Das Rentensystem verstehen: Einfuehrung in die politische Oekonomie der Alterssicherung* [Understanding the pension system: Introduction into the political economy of old-age provision] (2nd ed.). Wochenschau Verlag.

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, *3*, 33–67. https://doi.org/10.1007/s13194-012-0056-8

Leiser, D., & Shemesh, Y. (2018). *How we misunderstand economics and why it matters: The psychology of bias, distortion and conspiracy*. Routledge.

Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, *42*(3), 847–862. https://doi.org/10.3758/BRM.42.3.847

Mayring, P. (2015). *Qualitative Inhaltsanalyse: Grundlagen und Techniken* [Qualitative content analysis: Basics and techniques] (12th ed.). Beltz.

Meadows, D. H. (2008). *Thinking in systems: A primer* (D. Wright, Ed.). Chelsea Green Publishing.

Mislevy, R. J., & Riconscente, M. M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology* (PADI Technical Report Series, 9). SRI International and University of Maryland.

Moosbrugger, H., & Kelava, A. (Eds.). (2020). *Testtheorie und Fragebogenkonstruktion* [Test theory and questionnaire construction] (3rd ed.). Springer. https://doi.org/10.1007/978-3-662-61532-4

Moosbrugger, H., Schermelleh-Engel, K., Gaede, J. C., & Kelava, A. (2020). Testtheorien im Ueberblick [Test theories at a glance]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* [Test theory and questionnaire construction] (pp. 251–273). Springer.

Pruefer, P., & Rexroth, M. (2005). Kognitive Interviews [Cognitive interviews]. *GESIS-How-to*, *15*, 1–21. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-201470

Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.

Toulmin, S. E. (2003). *The uses of argument* (Updated ed.). Cambridge University Press.

Wetzel, E., & Boehnke, J. R. (2017). Differential Item Functioning. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *Encyclopedia of Personality and Individual Differences* (pp. 1–5). Springer International Publishing.

7 Appendix

The following appendix includes an English version of the GePSUS. It was translated to increase the traceability of the description regarding the development and validation process in Baginski (2025). However, it is not advisable to use or adapt the English version of the GePSUS without further validation studies.

GePSUS (German pension system understanding scales)

General instructions to the respondents:
In the following section, I will ask you some questions about different forms of old-age provision in Germany. If you do not know the answer to a question, please let me know. But try to think about the questions first, please.

Instructions to the interviewee:
Do not read out loud the response options "Don't know" and "No statement".
Choose "Don't know" if a respondent does not know the answer and "No statement" if the respondent refuses to answer.

Scale 1: Understanding the statutory pension insurance (SPI)
Instructions to the respondents:
The first questions are about the statutory pension insurance.

Scale 1 (SPI) – Part 1/3:
Instructions to the respondents:
Now we are interested in the contributions to the statutory pension insurance. Please indicate whether you think the following statements apply or do not apply.
[Not randomised]

No.	Wording	Response options (correct ones underlined)
1.1	In Germany all persons, whether they are self-employed or not, are obliged to pay contributions to the statutory pension insurance.	1 Applies 2 Does not apply 3 Don't know 999 No statement
1.2	In Germany, contributions to the statutory pension insurance are paid in full by employees.	1 Applies 2 Does not apply 3 Don't know 999 No statement
1.3	In Germany, civil servants are obliged to pay contributions to the statutory pension insurance.	1 Applies 2 Does not apply 3 Don't know 999 No statement

Scale 1 (SPI) – Part 2/3:
Instructions to the respondents:
And what do you think happens to the contributions paid into the statutory pension insurance? Please state again whether you think these statements apply or do not apply.
[Not randomised]

2.1	The contributions are saved by the statutory pension insurance in a secured account for every insured person.	1 2 3 999	Applies Does not apply Don't know No statement
2.2	The contributions are invested by the statutory pension insurance on the capital market.	1 2 3 999	Applies Does not apply Don't know No statement
2.3	The contributions are used to achieve a more equitable distribution between retirees with lower income and those with higher income.	1 2 3 999	Applies Does not apply Don't know No statement
2.4	The contributions are used to finance the pension payments to the persons who are currently retired.	1 2 3 999	Applies Does not apply Don't know No statement
Scale 1 (SPI) – Part 3/3: The following part is about the paid-out pension amount. Please tell me whether you think that the following factors have a large influence on how much pension income someone receives from the statutory pension insurance. Please answer Yes if you think they have a large influence and No if this is not the case. [Randomised]			
3.1	The way in which the 'Deutsche Rentenversicherung' (i.e., the German statutory pension agency) invests the money has a large influence on how much pension income someone receives from the statutory pension insurance.	1 2 3 999	Yes No Don't know No statement
3.2	The development of interest rates on the capital market has a large influence on how much pension income someone receives from the statutory pension insurance.	1 2 3 999	Yes No Don't know No statement
3.3	The amount of the last income before retirement of the insured person has a large influence on how much pension income someone receives from the statutory pension insurance.	1 2 3 999	Yes No Don't know No statement
3.4	The administration costs of the 'Deutsche Rentenversicherung' have a large influence on how much pension income someone receives from the statutory pension insurance.	1 2 3 999	Yes No Don't know No statement
Scale 2: Understanding a private pension insurance (PPI)			
Scale 2 (PPI) – Part 1/2 Instructions to the respondents: We now turn to a private pension insurance as another form of old-age provision in Germany. Please indicate whether you think the following statements apply or do not apply. [Not randomised]			
4.1	With a private pension insurance, a certain amount of money is paid to an insurer, which can	1 2	Applies Does not apply

	be converted into a monthly pension at the start of the retirement period.	3 Don't know 999 No statement
4.2	A private pension insurance is obligatory for self-employed persons if they do not pay into the statutory pension insurance.	1 Applies 2 Does not apply 3 Don't know 999 No statement
4.3	In order to receive a planned amount of a pension from a private pension insurance, the following applies: The longer you pay, the lower the amount that you have to pay in monthly.	1 Applies 2 Does not apply 3 Don't know 999 No statement
Scale 2 (PPI) – Part 2/2 And what do you think has a large influence on how much pension income someone receives from a private pension insurance? I will read out some factors again. Please answer Yes if you think they have a large influence and No if this is not the case. [Randomised]		
5.1	The way in which the insurer invests the money has a large influence on how much pension income someone receives from a private pension insurance.	1 Yes 2 No 3 Don't know 999 No statement
5.2	The development of interest rates on the capital market has a large influence on how much pension income someone receives from a private pension insurance.	1 Yes 2 No 3 Don't know 999 No statement
5.3	The amount of the last income before retirement of the insured person has a large influence on how much pension someone receives from a private pension insurance.	1 Yes 2 No 3 Don't know 999 No statement
5.4	The administration costs of the respective insurer have a large influence on how much pension income someone receives from a private pension insurance.	1 Yes 2 No 3 Don't know 999 No statement
Scale 3 (Factors) – Part 1/1 Certain long-term changes can affect the stability of the entire system of old-age provision. I will now read out four of such changes to you. If nothing else changes, how does the respective change affect the stability of the statutory pension insurance or private pension insurance? [Randomised]		
6.1	Does a decrease of the interest rate have a larger long-term effect on the stability of the statutory pension insurance or on the stability of a private pension insurance or on both of about the same extent or on neither of them?	1 Larger on the statutory pension insurance 2 Larger on a private pension insurance 3 On both of about the same extend 4 On neither of them 5 Don't know 999 No statement

6.2	Does a rise in the unemployment rate have a larger long-term effect on the stability of the statutory pension insurance or on the stability of a private pension insurance or on both of about the same extent or on neither of them?	1 Larger on the statutory pension insurance 2 Larger on a private pension insurance 3 On both of about the same extend 4 On neither of them 5 Don't know 999 No statement
6.3	Does a rise in the inflation rate have a larger long-term effect on the stability of the statutory pension insurance or on the stability of a private pension insurance or on both of about the same extent or on neither of them?	1 Larger on the statutory pension insurance 2 Larger on a private pension insurance 3 On both of about the same extend 4 On neither of them 5 Don't know 999 No statement

8 Contact

For questions regarding the test documentation:

Dr. Ronja Baginski ronja.baginski@uni-mannheim.de

For questions regarding the CHF Open Access Repository:

Prof. Dr. Carmela Aprea carmela.aprea@uni-mannheim.de