Measuring Hygiene Competence: A Vignette-Based Situational Judgement Test

Susanne Katharina Heininger (susanne.heininger@tum.de)
Technische Universitat Munchen Fakultat fur Medizin https://orcid.org/0000-0001-7678-8533

Maria Baumgartner
Klinikum rechts der Isar der Technischen Universitat Munchen, TUM Medical Education Center

Fabian Zehner
DIPF - Leibniz-Institut für Bildungsforschung und Bildungsinformation

Rainer Burgkart
Klinikum rechts der Isar der Technischen Universität München, Klinik und Poliklinik für Orthopädie und Sportorthopädie

Nina Söllner
Klinikum rechts der Isar der Technischen Universität München, TUM Medical Education Center

Pascal O. Berberat
Klinikum rechts der Isar der Technischen Universität München, TUM Medical Education Center

Martin Gartmeier
Klinikum rechts der Isar der Technischen Universität München, TUM Medical Education Center

---

Research article

Keywords: situational judgement test, hygiene, competence, assessment, item-response theory

Posted Date: May 29th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-30686/v1

Version of Record: A version of this preprint was published at BMC Medical Education on July 30th, 2021. See the published version at https://doi.org/10.1186/s12909-021-02829-y.

Measuring Hygiene Competence:

A Vignette-Based Situational Judgement Test

Susanne Heininger[1], *susanne.heininger@tum.de*

Maria Baumgartner[1], *maria-baumgartner@arcor.de*

Fabian Zehner[2], fabian.zehner@dipf.de

Rainer Burgkart[3], *rainer.burgkart@tum.de*

Nina Söllner[1], *n.soellner@tum.de*

Pascal O. Berberat[1], *berberat@tum.de*

Martin Gartmeier[1], *martin.gartmeier@tum.de*

[1]TUM Medical Education Center, Fakultät für Medizin, Klinikum rechts der Isar, TU München
[2] DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
[3]Klinik und Poliklinik für Orthopädie und Sportorthopädie, Fakultät für Medizin, Klinikum rechts der Isar, TU München

Correspondence:

Dr. Susanne Katharina Heininger

Technical University of Munich

University Hospital *Klinikum rechts der Isar*

TUM Medical Education Center

Ismaninger Straße 22, D-81675 München

Tel. +49 89 4140 9278

Mail: susanne.heininger@tum.de

Measuring Hygiene Competence:

A Vignette-Based Situational Judgement Test

Abstract

Background:

Since the onset of the Corona pandemic at the beginning of 2020, the extreme importance of hygiene has once again become very clear. In the medical context, it is not easy to find suitable test formats to assess the competencies involved in "working hygienically". Pre-existing test formats usually use self-report questionnaires, which are suboptimal for this purpose.

Methods:

We designed a Situational Judgement Test (SJT) to assess hygiene competence. The SJT incorporates reliable measures and is a good predictor of job performance. The test consists of twenty pictures (items) presenting *only one unambiguous* hygiene lapse. Test respondents were asked (1) to note down whether they recognized a problem in the picture in respect of hygiene guidelines and (2), if yes, to describe the problem in a short verbal response. The sample comprised n = 149 health care professionals (79.1% female; age: M = 26.7 years, SD = 7.3 years) working as clinicians or nurses. The written responses were rated by two independent raters with high agreement ($\alpha > .80$), indicating high reliability of the measurement. We then used Item Response Theory (IRT) for data analysis.

Results:

The reported IRT analyses show that the test is suitable to assess hygiene competence and that it is possible to distinguish between persons demonstrating different levels of ability for seventeen of the twenty items/pictures), expecially for the range of low to medium person abilities.

Conclusions:

The test in its present form can be used to assess the hygiene competence of medical students, medical doctors, nurses and trainee nurses in cross-sectional measurements. In order to broaden the difficulty spectrum of the current test, additional test items with higher difficulty should be developed. The Situational Judgement Test designed to assess hygiene competence can be helpful in *testing* and *teaching* the ability of working hygienically.

Keywords: situational judgement test; hygiene; competence; assessment; item-response theory;

68 *Background*

69 Nosocomial infections are a serious challenge in modern patient care [1] and have recently been
70 subject of intense research [2]. In order to prevent hospital-acquired infections (HAI) and improve
71 patient and staff safety, hygiene is essential. In pandemics like COVID-19 in the year 2020, the
72 prevention of infection through good hygiene is crucial. Some pre-existing forms of intervention are
73 designed to increase awareness and improve attitudes towards hygiene in daily clinical work [3–5].
74 While there are numerous single studies (e.g. [6], [7, 8] and a number of meta-analyses [9, 10] on the
75 issue of hygiene, these mostly focus on interventions for improving *hand* hygiene. A massive
76 international campaign for the prevention of infection, the "My Five Moments of Hand Hygiene"-
77 program [11], also puts the emphasis on hand hygiene. We argue, however, that the focus on hand
78 hygiene is rather narrow because health care staff draw upon a more differentiated repertoire of
79 cognitive resources in order to adhere to hygiene standards in daily work. This argument is supported
80 by the *WHO Guidelines on Core Components of Infection Prevention and Control (IPC) Programmes*
81 [12] . These guidelines, based on the *WHO Core Components for Infection Prevention and Control*
82 *Report*, were extended in 2016, highlighting additional areas for preventing infections in health care
83 facilities worldwide, especially in acute health care facilities: personnel and facility resources,
84 including workload, staffing, materials and equipment (guideline recommendations numbers seven
85 and eight). This points up the need to develop a more differentiated concept of hygiene competence
86 – which also requires more sophisticated measurement strategies.

87 In a recent article, Gartmeier et al. [13] proposed a multidimensional model of *hygiene competence*
88 consisting of three dimensions, *knowledge*, *skills* and *attitudes*. They argue that *knowledge* about
89 why hygiene is important and how it can be maintained in specific clinical situations is necessary, but
90 not sufficient. Moreover, health care staff need specific *skills* in order to apply their knowledge in
91 hygienic patient care. Finally, specific *attitudes* are required, e.g. maintaining hygienic working
92 procedures despite extreme time pressures and understaffing. What this model proposes (especially
93 compared to more general models such as the "My Five Moments of Hand Hygiene" [11]) is that in

order to uphold hygiene across a multitude of particular health care tasks (e.g., providing artificial respiration or replacing a patient's bladder catheter), specific procedures are performed which require synergy between all three dimensions of hygiene [13] described above.

In order to improve hygiene competence more effectively in the future, we argue that psychometrically promising instruments should be designed to assess this competence. The approach to measurement should take account of the fact that health care staff manage to synthesize different resources – knowledge, skills and attitudes – in order to maintain hygiene in their daily work practice. We propose that a *Situational Judgement Test* (SJT) requiring test respondents to identify the presence or absence of hygiene problems in images of clinical situations is a promising approach in this respect. We elaborate this conjecture in what follows.

Assessment of hygiene competence

Currently, researchers primarily use self-report questionnaires to assess hygiene-related constructs with regard to attitudes and practices (see e.g. [6] [14]). Evidence shows, however, that such instruments are suboptimal in measuring competencies [15], partly because health care staff systematically overestimate the number of hygiene-related behaviors they perform in clinical practice [16]. More objective measures should therefore be developed to assess hygiene competence. In order to measure competencies, direct observations or simulations are optimal but are resource-intensive and cannot be easily standardized [17, 18]. With regard to hand hygiene, different observational tools are available [19–21], but no current instrument adopts the more comprehensive measurement approach suggested by the above mentioned model of hygiene competence [13]. Drawing upon this model, we propose that a SJT is a promising, time- and cost-efficient method of assessment. SJTs require respondents to make knowledge-based judgements of scenarios displayed as short texts, pictures or videos [22] [23]. Properly designed SJTs yield reliable measures, they are good predictors of job performance and are well accepted by test respondents [23]. Moreover, SJTs allow assessments to be conducted in a standardized way – which is less of a

drain on scarce resources and time [22, 23]. Such instruments are hence increasingly used in the medical context, for instance, in context of medical admission procedures [22–24].

In the Methods section below, we describe our design of an SJT intended to measure hygiene competence, or, more specifically, the dimensions *knowledge* and *practices/skills.* This test requires respondents to make knowledge-based judgements of still-images of common clinical work procedures and situations. Respondents have to apply their knowledge in order to judge whether the clinical practices shown in the images are carried out in compliance with hygiene standards. This assessment strategy relates to the second level in Miller's widely used prismatic model of clinical competence [17], which represents "knows how" as the (cognitive) application of knowledge. We will report a pilot study with an initial version of this SJT, focusing upon two research questions: (RQ1) Does the SJT comply with competence test quality requirements described in context of the Item-Response-Theory (IRT)? (RQ2) Is it possible to cover a broad spectrum of item difficulty / person ability regarding hygiene competence by means of the SJT?

Methods

SJT design and measurement strategy

Our SJT is designed to assess hygiene competence and consists of twenty picture vignettes (see Figure 1 for a sample item). Each vignette shows at least one health care provider (nurse [f/m] or clinician) and a patient (where appropriate) in a clinical situation in which hygiene is a relevant issue. The test items were constructed in the following way: Every item shows *only one unambiguous* fault (two items show no fault). The situations displayed (medical equipment and rooms) are *authentic reconstructions* of clinical situations taken in different rooms (OR, patient room, treatment room, intensive care room) of the Medical Training Center (MTC) of the München rechts der Isar (MRI) university hospital. All situations shown in the test items focus on *different* actions of clinicians and nurses (f/m; see Table 1). The reason for the interdisciplinary setup lies in the fact that hygiene

144 competence often needs to be applied in collaborative settings. Hygiene fails when two professions
145 fail to achieve inter-professional alignment. A thorough knowledge of hygiene is therefore important
146 regardless of profession. As described above, these items refer to the hygiene model dimensions
147 *knowledge* and *skills* [13]. The *attitudes*-dimension of the competence model could not directly be
148 represented in the test. All persons depicted were non-professional actors from whom full consent
149 was obtained. Materials, rooms, equipment and procedures were pictured as realistically as possible
150 (e.g. picture vignettes showing a situation in OR were taken in a simulation OR with original
151 materials). All situations displayed were conceptualized in close cooperation with and reviewed by
152 nurses and medical doctors with extensive clinical experience to ensure the most realistic portrayal
153 of situations in the pictures.

154 *Insert Figure 1 about here*

155 The generated picture vignettes were compiled into a paper-pencil questionnaire. For each vignette,
156 respondents were asked (1) to note down whether they recognize a problem in the picture regarding
157 hygiene guidelines (dichotomous: yes/no) and (2), if yes, to describe the problem in a short verbal
158 response. For each item, an ideal solution was developed based on scientific literature and national
159 clinical hygiene guidelines. All picture vignettes and solutions were reviewed in cooperation with the
160 group of local hygiene experts (who were advanced clinicians, nurses [f/m] and specialists from the
161 local hygiene department). Only *completely* correct answers were credited if (1) the respondent
162 recognized an incorrect (unhygienic) action in the picture *and* (2) described the hygiene lapse
163 correctly (matching to the model solution). In contrast, respondents were not credited if they just
164 recognized the incorrect (unhygienic) action, but could not point out what exactly was wrong (e.g.
165 sterile gloves missing). In the present study, three raters (two pairs: raters A+B, raters B+C) evaluated
166 all given answers and took a pass-fail-decision (dichotomous) based on the model solution.
167 Krippendorff's alpha [25] was calculated to measure the inter-rater reliability of the rating procedure.

168

169 *Table 1. Content of the test items*

Item	Content / Issue	Hygiene problem
1	Incorrect wearing of surgical face mask	yes
2	Incorrect hand posture during pre-surgery dressing	yes
3	Nurse incorrectly standing next to a sterile table	yes
4	Wearing of personal items (wristwatch) at bedside	yes
5	Personal sluice (incorrect changing of personal to work clothing)	Yes
6	Non-sterile suctioning of mucus	Yes
7	Putting blood sampling set down on patient's bedside	Yes
8	Personal hygiene – nurse's hair touching patients bed	Yes
9	Sterile hand gloves being transported in nurse's dress pocket	Yes
10	Personal hygiene – tie touching patient's bed blanket	Yes
11	Personal hygiene (artificial fingernails) in clinical practice	Yes
12	Reaching into care trolley drawer with contaminated hand gloves	Yes
13	Using care trolley in isolated patient room	Yes
14	Wearing medical work clothes in public areas	Yes
15	No-touch-technique	No
16	Hygienically transporting medical file on patient bed	No
17	Typing on computer keyboard wearing sterile hand gloves	Yes
18	Carrying infusion bottle in nurse's dress pocket	Yes
19	Opening sterile syringe packaging	Yes
20	Wearing personal items (jewelry) in clinical practice	Yes

170

171 *Sample*

172 The sample comprised *n* = 149 health care professionals (79.1% female; age: *M* = 26.7 years, *SD* = 7.3

173 years) working as clinicians or nurses (f/m; see Table 2 for overview).

174 *Table 2. Sample professions and professional experiences*

Profession	**Frequency** *in %*	**Professional Experience in Years**	
		M (*SD*)	Freq. *in %*
Medical Student	40.8	5.6 (.85)	
Medical Doctor	4.8	7.0 (8.49)	
Trainee Nurse[a] (f/m)	34.7	2.2 (1.00)	First year: 41.7 Third year: 58.3
Nurse (f/m)	19.0	14.3 (13.81)	
Trainee operating room technician (ORT)[b]	0.7	1.0 (0)	First year: 100

175 [a] In Germany, nursing education is a vocational college training, which is not rooted at a university. Trainee nurses spent
176 three years at a vocational school for theoretical and practical education (2,100h). Additionally, trainee nurses serve 2,500
177 hrs at a teaching hospital for practical training.
178 [b] In Germany, ORT education is based at a vocational college as well. The theoretical education covers 1,600 hrs, the
179 practical training includes 3,000 hrs at a teaching hospital.

180

181 *Analysis*

182 We used Item Response Theory (IRT) for data analysis. The IRT framework provides a family of

183 probabilistic models, which allow for simultaneous estimation of respondent ability and item

184 difficulty [26]. This way, one can model "the interaction between an individual item and an individual

185 examinee" [27] by using the Rasch Model, the most restrictive IRT model with the best measurement

186 properties. Only if a test has been shown to produce data conforming to the Rasch Model is it fair to

187 use the sum score. This model can be used to estimate unidimensional latent abilities [26]. The

188 model assumes that the probability $P(X_{v,i})$ of a person *v* giving a certain answer *X* for an item *i*

189 depends on two factors: (1) the item difficulty, σ_i, and (2) the person's ability θ_v [28], which are

190 estimated simultaneously on the same scale. This means, the numbers for person abilities and item

difficulties can be directly compared. If a person with ability θ_v = 1.1 responds to an item with difficulty σ_i = 1.1, the probability of a correct response is 50 percent. But if the difference between ability and difficulty is largely positive or negative, the probability of a correct response gets close to 100 percent or 0 percent, respectively. Negative parameter values indicate simple items or low ability; positive values refer to difficult items or high ability [29]. Mathematically, these relationships are modelled by the following equation:

$$PXi,=1\ \ \sigma i,\theta v= e(\theta v\text{-}\sigma i)1+e(\theta v\text{-}\sigma i)$$

In order to test whether collected data are compliant with the Rasch Model, it is crucial to confirm the model's assumption of *specific objectivity* [26]. This means that different subpopulations in a sample should not differ significantly regarding item difficulties. To measure this, is current state-of-the-art involves splitting a sample into subsamples and comparing the resulting item difficulties using the *Andersen Likelihood Ratio Test* [30]. Since the null hypothesis assumes no differences, it is common to use three split criteria, and to apply a Type-I-risk of α = .01, so this risk inflates to no more than α = .05 across all comparisons [26]. In this study, we used the split criteria *raw score median* (comparing two achievement groups), *gender* and *age median*. IRT analyses were carried out using the R packages *eRm* and *PP* [31–33].

Results

Inter-rater-reliability

We calculated Krippendorff's alpha for inter-rater reliablity and found a strong agreement (α ≥ .80; see Table 3).

Table 3. Krippendorff's alpha for rater sets

Rater Group	**Krippendorff's alpha**
Raters A+B	.81
Raters B+C	.86

Note. Rater A: student rater; rater B/C: research associates

IRT Scaling (research questions one and two)

Based on Andersen's Likelihood Ratio Test, three of the twenty items had to be excluded in order to reach sufficient model fit. Using the remaining seventeen items, Rasch Model conformity could be assumed for the instrument – see test statistics displayed in Table 4. Fisher's *z*-Test indicated which item had to be excluded. Figure 1 shows the resulting item difficulty parameters and person parameter distributions. An item's measurement is most informative if its difficulty is about equal to a person's ability.

Table 4. Rasch Model Conformity Achieved for the Final Scale with 17 Items

Split Criterion	**LRT χ^2**	df	$\chi^2_{df,\alpha=.01}$	*p*
High / Low Achievement	28.48	16	32.00	.028
Female / Male	31.81	16	32.00	.011
Younger / Older	31.87	16	32.00	.010

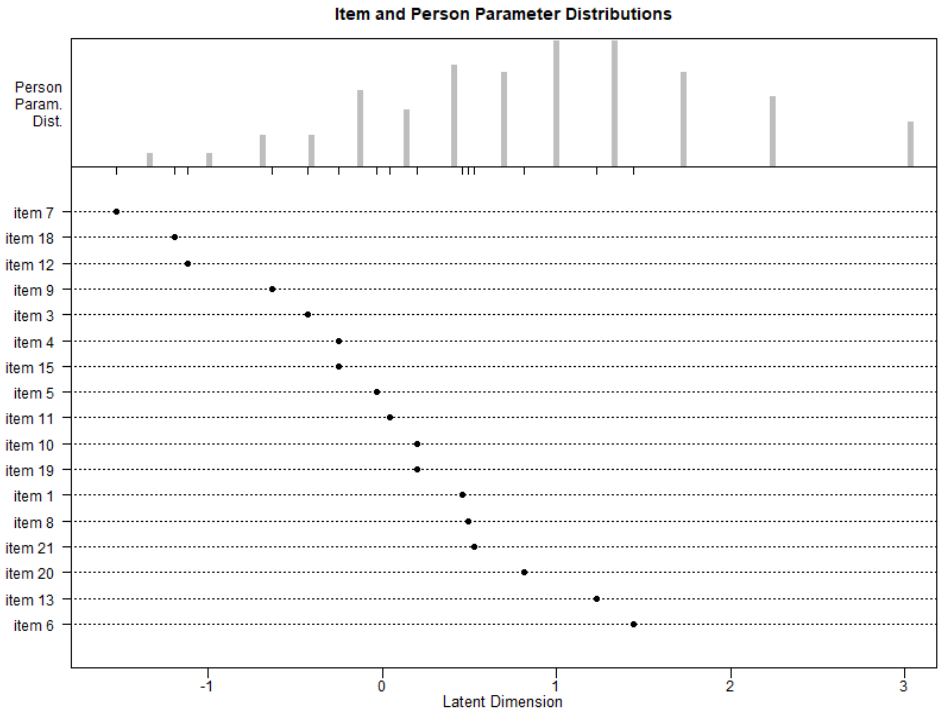

Figure 1. Person ability (top) and item (bottom) parameter distribution for the final instrument

As is apparent, our test items cover a difficulty spectrum from low to medium; no item showed a high degree of difficulty. Some test respondents, however, showed higher ability, meaning they were seemingly able to solve all test items (cf. upper section of Figure 1).

Discussion

In this paper, we have described a novel, time and cost-efficient, yet psychometrically promising method of assessing hygiene competence based upon a recent theoretical model [13]. The developed SJT, originally consisting of twenty picture items, offers a feasible way of assessing hygiene competence. The responses given by the study participants were rated by two independent raters with high agreement ($\alpha > .80$), indicating the high reliability of the measurement. Further, the reported IRT analyses show that the test is suitable to assess hygiene competence and that it is possible to distinguish between persons with different levels of ability. After exclusion of three items, the remaining seventeen items showed a good measurement quality (RQ1). Further, the results showed that our test items cover the range of low to medium person ability regarding hygiene competence. The current results show that the initial test version does not contain items with sufficient difficulty to discriminate for very high abilities (i.e., highly able test respondents would answer all items correctly, RQ2). Nevertheless, because we did not observe a strong ceiling effect, the test in its present form can be used to assess hygiene competence of medical students, medical doctors, nurses and trainee nurses in cross-sectional measurements.

A general conclusion to be drawn from the present results is that the described measurement strategy offers a promising means of estimating professionals' hygiene competence. Nevertheless, some critical remarks have to be made regarding the test in its current version: First, the test covers a broad range of clinical situations, some of which may be regarded as special cases (e.g. items 1, 2, 3 and 5 showing situations in the OR). For this reason, it could be argued that these items are very difficult to answer for respondents who have no or only very limited experience regarding this very

specific area of health care. However, as is apparent from Table 1, the operating-room items 1, 3 and 5 all had medium difficulty levels. Thus, these outcomes *do not suggest* that respondents had substantial difficulties in answering these items because they were not familiar with these situations.

In order to further develop the current instrument, several perspectives are recommended: First, we suggest a data-based strategy to determine what is measured by the current instrument. To achieve this, correlations with data collected with existing hygiene-related measures (e.g. [6] [14]) should be examined. In addition, correlations with performance in simulated situations or with observations in clinical practice would be useful. In order to broaden the difficulty spectrum of the current test, additional test items with higher difficulty could be developed. This can be achieved in terms of content by using more specific situations, e.g. in the OR, or with special treatment situations. An alternative solution would be to use a more complex item format, such as short videos instead of picture vignettes. Hygiene lapses would be viewable only for a short interval, making them more authentic and closer to real-world performance. The video format would provide an approach more consistent with reality, but would also require investigating the attention processes of the test respondents. Furthermore, the video format can be seen as a connection between a static photo and the fast-moving reality of real-world health care scenarios. It would allow complex reality to be presented with the options of stopping or slowing the playback. Video items could be used in *testing* hygiene competence as well as in *teaching* hygiene competence.

Conclusions

Since the onset of the Corona pandemic at the beginning of 2020, the extreme importance of hygiene is clear to everyone in the health care sector and beyond. Teaching hygiene to medical students may thus receive even greater attention than before. For teaching purposes, it is not always easy to find suitable test formats that make it possible to test newly acquired skills. It is therefore even more important to be able to test the results of the competencies taught. This Situational

276 Judgment Test, which is designed to assess hygiene competence, may help to address the urgent
277 need to teach and test the ability to work hygienically in the medical context.

278

279 ***Abbreviations***

280 COVID-19	Coronavirus disease (SARS-CoV-2)
281 IPC	Infection prevention and control. Infection prevention and control (*IPC*) of WHO is a 282 scientific approach and practical solution designed to prevent harm caused by 283 infection to patients and health workers.
284 IRT	Item Response Theory
285 MRI	University hospital rechts der Isar (located to the right of the Isar river)
286 MTC	Medical training center
287 OR	Operating room
288 RQ	Research question
289 SJT	Situational judgement test
290 WHO	World health organization

291

292

293 ***Declarations***

294 *Ethics approval and consent to participate*

295 All procedures performed in studies involving human participants were in accordance with the
296 ethical standards of the institutional and/or national research committee and with the 1964 Helsinki
297 declaration and its later amendments or comparable ethical standards. The data acquisition

procedure was approved by the local ethics commission (No. 168/18 S, Ethics Commission, Klinikum rechts der Isar der Technischen Universität, München).

Consent for publication

Not applicable.

Availability of data and materials

All data and materials are available from the manuscript or from the corresponding author upon request.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

All authors have contributed sufficiently to the project to be included as authors. MG POB RB MB and SH designed the study. MG POB NS RB MB and SH coordinated the study and the data acquisition. FZ MG NS MB and SH performed the statistical analyses. SH MG FZ POB drafted the manuscript. All authors revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements

Many thanks to all medical students who participated in this study and to the many voluntary supporters. No internal or external sources of funding were used to support this work.

Authors' information

TUM Medical Education Center, Fakultät für Medizin, Klinikum rechts der Isar, TU München

Susanne Heininger, Maria Baumgartner, Nina Söllner, Pascal O. Berberat, Martin Gartmeier

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Fabian Zehner

Klinik und Poliklinik für Orthopädie und Sportorthopädie, Fakultät für Medizin, Klinikum rechts der Isar, TU München

Rainer Burgkart

References

1. Khan HA, Baig FK, Mehboob R. Nosocomial infections: Epidemiology, prevention, control and surveillance. Asian Pacific Journal of Tropical Biomedicine. 2017;7:478–82. doi:10.1016/j.apjtb.2017.01.019.
2. World Health Organization. WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge. Clean Care is Safer Care. Geneva: WORLD HEALTH ORGANIZATION; 2009.
3. Cresswell P, Monrouxe LV. 'And you'll suddenly realise 'I've not washed my hands': medical students', junior doctors' and medical educators' narratives of hygiene behaviours. BMJ Open. 2018;8:e018156. doi:10.1136/bmjopen-2017-018156.
4. Gilmartin H, Saint S, Rogers M, Winter S, Snyder A, Quinn M, Chopra V. Pilot randomised controlled trial to improve hand hygiene through mindful moments. BMJ Qual Saf. 2018;27:799–806. doi:10.1136/bmjqs-2017-007359.
5. Pittet D, Hugonnet S, Harbarth S, Mourouga P, Sauvan V, Touveneau S, Perneger TV. Effectiveness of a hospital-wide programme to improve compliance with hand hygiene.: Infection Control Programme. Lancet. 2000;356:1307–12.
6. Graf K, Chaberny IF, Vonberg R-P. Beliefs about hand hygiene: a survey in medical students in their first clinical year. Am J Infect Control. 2011;39:885–8. doi:10.1016/j.ajic.2010.08.025.
7. Fichtner A, Haupt E, Karwath T, Wullenk K, Pöhlmann C, Jatzwauk L. A single standardized practical training for surgical scrubbing according to EN1500: effect quantification, value of the standardized method and comparison with clinical reference groups. GMS Z Med Ausbild. 2013;30:Doc24. doi:10.3205/zma000867.
8. Fuller C, Michie S, Savage J, McAteer J, Besser S, Charlett A, et al. The Feedback Intervention Trial (FIT)--improving hand-hygiene compliance in UK healthcare workers: a stepped wedge cluster randomised controlled trial. PLoS ONE. 2012;7:e41617. doi:10.1371/journal.pone.0041617.
9. Erasmus V, Daha TJ, Brug H, Richardus JH, Behrendt MD, Vos MC, van Beeck EF. Systematic review of studies on compliance with hand hygiene guidelines in hospital care. Infect Control Hosp Epidemiol. 2010;31:283–94. doi:10.1086/650451.
10. Kingston L, O'Connell NH, Dunne CP. Hand hygiene-related clinical trials reported since 2010: a systematic review. J Hosp Infect. 2016;92:309–20. doi:10.1016/j.jhin.2015.11.012.

357 11. Sax H, Allegranzi B, Uçkay I, Larson E, Boyce J, Pittet D. 'My five moments for hand hygiene': a
358 user-centred design approach to understand, train, monitor and report hand hygiene. J Hosp
359 Infect. 2007;67:9–21. doi:10.1016/j.jhin.2007.06.004.

360 12. World Health Organization. Guidelines on core components of infection prevention and control
361 programmes at the national and acute health care facility level. WORLD HEALTH ORGANIZATION;
362 2016.

363 13. Gartmeier M, Baumgartner M, Burgkart R, Heininger SK, Berberat PO. Why Hand Hygiene is Not
364 Sufficient: Modeling Hygiene Competence of Clinical Staff as a Basis for its Development and
365 Assessment. GMS Journal Med Ed. 2019;36.

366 14. Richter A, Chaberny IF, Surikow A, Schock B. Hygiene in medical education – Increasing patient
367 safety through the implementation of practical training in infection prevention. GMS Journal
368 Med Ed. 2019;36:1–16. doi:10.3205/zma001223.

369 15. Hartig J, Klieme E, Leutner D, editors. Assessment of Competencies in Educational Contexts. 1st
370 ed. Göttingen: Hogrefe Publishing; 2008.

371 16. Jenner EA, Fletcher BC, Watson P, Jones FA, Miller L, Scott GM. Discrepancy between self-
372 reported and observed hand hygiene behaviour in healthcare professionals. J Hosp Infect.
373 2006;63:418–22. doi:10.1016/j.jhin.2006.03.012.

374 17. Mehay R, Burns R. Miller's Pyramid of Clinical Competence. In: Mehay R, editor. The essential
375 handbook for GP training and education. London: Radcliffe Publishing Ltd; 2012. p. 414.

376 18. Heininger SK. Die Qualität bildungswissenschaftlicher Evidenz erkennen. Wiesbaden: Springer
377 Fachmedien Wiesbaden; 2019.

378 19. McAteer J, Stone S, Fuller C, Charlett A, Cookson B, Slade R, et al. Development of an
379 observational measure of healthcare worker hand-hygiene behaviour: the hand-hygiene
380 observation tool (HHOT). Journal of Hospital Infection. 2008;68:222–9.
381 doi:10.1016/j.jhin.2007.12.009.

382 20. McCalla S, Reilly M, Thomas R, McSpedon-Rai D. An automated hand hygiene compliance system
383 is associated with improved monitoring of hand hygiene. Am J Infect Control. 2017;45:492–7.
384 doi:10.1016/j.ajic.2016.12.015.

385 21. Sax H, Allegranzi B, Chraïti M-N, Boyce J, Larson E, Pittet D. The World Health Organization hand
386 hygiene observation method. Am J Infect Control. 2009;37:827–34.
387 doi:10.1016/j.ajic.2009.07.003.

388 22. Whetzel DL, McDaniel MA. Situational judgment tests: An overview of current research. Human
389 Resource Management Review. 2009;19:188–202. doi:10.1016/j.hrmr.2009.03.007.

390 23. Patterson F, Zibarras L, Ashworth V. Situational judgement tests in medical education and
391 training: Research, theory and practice: AMEE Guide No. 100. Med Teach. 2016;38:3–17.
392 doi:10.3109/0142159X.2015.1072619.

393 24. Kiessling C, Bauer J, Gartmeier M, Iblher P, Karsten G, Kiesewetter J, et al. Development and
394 validation of a computer-based situational judgement test to assess medical students'
395 communication skills in the field of shared decision making. Patient Educ Couns. 2016;99:1858–
396 64. doi:10.1016/j.pec.2016.06.006.

25. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Communication Methods and Measures. 2007;1:77–89.

26. Kubinger KD. Psychological Test Calibration Using the Rasch Model—Some Critical Suggestions on Traditional Approaches. International Journal of Testing. 2005;5:377–94. doi:10.1207/s15327574ijt0504_3.

27. van der Linden WJ, editor. Handbook of modern item response theory: Volume three: Applications. New York: Springer; 2010.

28. Geiser C, Eid M. Item-Response-Theorie. In: Wolf C, Best H, editors. Handbuch der sozialwissenschaftlichen Datenanalyse. 1st ed. Wiesbaden: VS Verlag für Sozialwissenschaften / Springer Fachmedien Wiesbaden GmbH Wiesbaden; 2010. p. 305–327. doi:10.1007/978-3-531-92038-2_14.

29. Bühner M. Einführung in die Test- und Fragebogenkonstruktion. 3rd ed. München: Pearson Studium; 2011.

30. Andersen EB. A goodness of fit test for the rasch model. Psychometrika. 1973;38:123–40. doi:10.1007/BF02291180.

31. Reif M. PP: Estimation of person parameters for the 1,2,3,4-PL model and the GPCM. R package version 0.6.1. 2017. https://github.com/manuelreif/PP.

32. Mair P, Hatzinger R, Maier M. J. eRm: Extended Rasch Modeling. 0.16-2. 2018. http://erm.r-forge.r-project.org/.

33. R Core Team. R: A language and environment for statistical computing. https://www.R-project.org/.