**Appendix A: Accountability of validity types and reliability**

Table A.1: Accountability of validity types and reliability in the gaming simulation design

|  |  |  |  |
| --- | --- | --- | --- |
| Validity concept | Description | Research design element | Applied in simulation development |
| Construct validity | Total aggregation of evidence based upon other validity types to support the interpretation of what a simulation reflects (Cronbach & Meehl, 1955; Polit & Beck, 2008) | Include feedback loops in the research design (Adams, 2015; Parshall & Guille, 2015). | See section 4.1 (Figure 1) for feedback loops. |
| Content validity | ‘The systematic examination of the content to determine whether the simulation represents the domain (Anastasi & Urbina, 1997, p. 114) | Calculate and evaluate the content validity ratio (Lawshe, 1975; Wilson, Pan, & Schumsky, 2012)\*  Statistical and qualitative evaluation of domain expert opinions (Sadeghi et al., 2022). | See section 5.3 and 6.3 for the qualitative and quantitative analysis and results. |
| Predictive validity | Predictive validity determines to which degree operationalisation provides accurate future outcomes (Lo & Meijer, 2013) | Create rules that resemble the rules of what is simulated as closely and accurately as possible (Dormans, 2011) | The simplified design was based upon a high level of precision. The outcome is shown in section 6.3.  Role-play is discussed in section 4.2. The rules of the gaming simulation are discussed in section 4.3. |
| Role-play results in accurate predictions or forecasts (Bolton, 2002; Green, 2002) |
| Internal validity | The extent to which the gaming simulation generates accurate and unambiguous conclusions from the results (De Vaus, 2010) | Simulation-based assessment based on participants’ actions  (Mislevy et al., 2016) Triangulation of research methods (Sari & Bogdan, 1992). | Participants by experience provided input for the simulated process.  An overview of the different methods is presented in Figure 1.  Rival explanations are discussed in section 7.  Participants by experience participated in every research method, see section 5. |
| The results support a claim about cause and effect (Shadish, Cook, & Campbell, 2002) | Address rival explanations (Stainton, Johnson, & Borodzicz, 2010).  Participants field experience is used in an applicable situation or process (Toulmin, 2003). |
| External validity | The extent to which results from this study are generalizable beyond this study (De Vaus, 2010; Feinstein & Cannon, 2002). | Examining multiple cases for replication (Stainton et al., 2010) | Multiple gaming simulation sessions were held at different companies. |
| Ecological validity | The extent to which what is simulated predicts similar outcomes in the real-world (Brunswik & Kamiya, 1953) | Consensus of experts regarding the content (Clauser, Margolis, & Clauser, 2016)  Participatory design (Spinuzzi, 2005) | The consensus of experts by experience is evaluated and analysed in section 6.  The applied participatory design is discussed in see section 2.1 and 5.1. |
| Reliability | The reliability of a gaming simulation is concerned with a consistent outcome after running repeated simulations (Berchtold, 2016). | Minimize error and bias of participants and researchers to obtain equal results when measurements (gaming simulation sessions) are repeated (Green, 2002). | Table 2 provides an overview of how consistent the outcomes are among the different simulation sessions. |

Legenda: \* Given the data it was not possible to calculate CVR in all cases and therefore not applied in this study, but included for the sake of completeness.

**References**

Adams, W. C. (2015). Conducting semi-structured interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (Vol. 492). Hoboken, New Jersey, USA: John Wiley & Sons, Inc.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*: Prentice Hall/Pearson Education.

Berchtold, A. (2016). Test–retest: agreement or reliability? *Methodological Innovations, 9*, 2059799116672875.

Bolton, G. E. (2002). Game theory’s role in role-playing. *International Journal of Forecasting, 18*(3), 353-358.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of'proximity'and of other Gestalt factors. *The American journal of psychology, 66*(1), 20-32.

Clauser, B. E., Margolis, M. J., & Clauser, J. C. (2016). Issues in simulation-based assessment. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 49-78). New York and London: Routledge Taylor & Francis Group.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281.

De Vaus, D. (2010). Research design in social research. *Research design in social research*, 1-296.

Dormans, J. (2011). Beyond iconic simulation. *Simulation & Gaming, 42*(5), 610-631.

Feinstein, A. H., & Cannon, H. M. (2002). Constructs of simulation evaluation. *Simulation & Gaming, 33*(4), 425-440.

Green, K. C. (2002). Forecasting decisions in conflict situations: a comparison of game theory, role-playing, and unaided judgement. *International Journal of Forecasting, 18*(3), 321-344.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel psychology, 28*(4), 563-575.

Lo, J. C., & Meijer, S. A. (2013). *Gaming simulation design for individual and team situation awareness.* Paper presented at the International Simulation and Gaming Association Conference.

Mislevy, R. J., Corrigan, S., Oranje, A., DiCerbo, K., Bauer, M. I., von Davier, A., & John, M. (2016). Psychometrics and game-based assessment. In F. Drasgow (Ed.), *Technology and testing: Improving educational and psychological measurement* (pp. 23-48). New York and London: Routledge Taylor & Francis Group.

Parshall, C. G., & Guille, R. A. (2015). Managing ongoing changes to the test: Agile strategies for continuous innovation. In *Technology and Testing* (pp. 1-22): Routledge.

Polit, D. F., & Beck, C. T. (2008). *Nursing research: Generating and assessing evidence for nursing practice*: Lippincott Williams & Wilkins.

Sadeghi, A. H., Peek, J. J., Max, S. A., Smit, L. L., Martina, B. G., Rosalia, R. A., . . . Mahtab, E. A. (2022). Virtual Reality Simulation Training for Cardiopulmonary Resuscitation After Cardiac Surgery: Face and Content Validity Study. *JMIR serious games, 10*(1), e30456.

Sari, K., & Bogdan, R. (1992). Qualitative research for education: An introduction to theory and methods.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*: Houghton, Mifflin and Company.

Spinuzzi, C. (2005). The methodology of participatory design. *Technical communication, 52*(2), 163-174.

Stainton, A. J., Johnson, J. E., & Borodzicz, E. P. (2010). Educational validity of business gaming simulation: A research methodology framework. *Simulation & Gaming, 41*(5), 705-723.

Toulmin, S. E. (2003). *The uses of argument*: Cambridge university press.

Wilson, F. R., Pan, W., & Schumsky, D. A. (2012). Recalculation of the critical values for Lawshe’s content validity ratio. *Measurement and evaluation in counseling and development, 45*(3), 197-210.