

Ameaças à Validade

O estudo sobre a viabilidade das bases de conhecimento do agente conversacional TOB-STT esteve sujeito a algumas ameaças. Conscientes disso, algumas medidas foram tomadas para mitigá-las:

Validade de conclusão: As conclusões corretas devem ser feitas a partir dos resultados. Assim, as medidas utilizadas precisam ser definidas de forma cuidadosa. Para fazer o cálculo da eficácia, este estudo teve como base o trabalho de Souza (2017), que considera as respostas emitidas pelos alunos em relação ao que era esperado para identificar a eficácia dos mesmos na atividade realizada no contexto de uma disciplina. Outra medida que foi usada no estudo é a qualidade das respostas emitidas pelo TOB-STT. Para apoiar a análise, utilizou-se procedimentos similares aos adotados no estudo de AbuShawar e Atwell (2016).

Validade interna: Como se trata de um estudo que abordou a habilidade dos estudantes em identificar defeitos em documentos, por meio de inspeção. Uma ameaça está relacionada ao treinamento que os alunos receberam sobre inspeção de documentos. Para evitar que um treinamento não adequado ocorresse, a professora que ministrou o treinamento sobre inspeção de documentos foi selecionada por possuir mais de 15 anos de experiência na área e já ter ministrado o tópico de interesse mais de cinco vezes. Além disso, como material de treinamento, utilizou-se módulos educacionais que foram produzidos utilizando atividades sistemáticas (i.e., um processo de desenvolvimento) e já haviam sido utilizados em pesquisas anteriores (BARBOSA et al. 2007).

Validade de constructo: Outro aspecto que teve que ser ponderado foi o viés de um único instrumento para coleta de dados. Durante o estudo de viabilidade, utilizou-se apenas um instrumento para coletar as opiniões dos participantes. Se fosse utilizado mais de um instrumento, como o método clínico de Piaget, é possível que resultados mais precisos teriam sido obtidos. Apesar disso, o instrumento utilizado para coletar a opinião dos participantes foi baseado em outros estudos que adotaram somente um questionário (e.g., Herpich et al. 2016 e Paschoal et al. 2018), uma vez que é um recurso menos custoso de ser utilizado e oferece uma visão geral sobre o fenômeno investigado.

Validade externa: Tendo em vista a generalização dos dados, este estudo utilizou estudantes de um curso de Computação, matriculados em uma disciplina específica de teste e inspeção de software.

Referências

ABUSHAWAR, B.; ATWELL, E. Usefulness, localizability, humanness, and language benefit: additional evaluation criteria for natural language dialogue systems. *International Journal of Speech Technology*, v. 19, n. 2, p. 373–383, 2016.

BARBOSA, E. F.; SOUZA, S. R. S. ; DOMINGUES, A. L. S. ; CHAN, A. ; HOHN, E. N. ; MALDONADO, J. C. . Uma Experiência no Ensino de Inspeção e Teste de Software. In: VI Simpósio Brasileiro de Qualidade de Software, 2007. p. 309-324.

HERPICH, F.; NUNES, F. B.; VOSS, G. B.; MEDINA, R. D. Three-dimensional virtual environment and npc: A perspective about intelligent agents ubiquitous. In: NETO, F. M. M.; SOUZA, R. de; GOMES, A. S. (Ed.). Handbook of Research on 3-D Virtual Environments and Hypermedia for Ubiquitous Learning. Hershey, EUA: IGI Global, 2016. cap. 21, p. 510–536.

PASCHOAL, L. N.; OLIVEIRA, M. M. de; CHICON, P. M. M. A chatterbot sensitive to student's context to help on software engineering education. In: XLIV Conferencia Latinoamericana de Informática. [S.l.: s.n.], 2018. p. 1–10.

SOUZA, D. M. Subsídios à integração de ferramentas de avaliação automática e sistemas de gerenciamento de aprendizagem. Tese (Doutorado) — Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2017.