

Acurácia da Autoavaliação Discente no Ensino Superior: Um Mapeamento Sistemático da Literatura

Title: Accuracy of Student Self-assessment in Higher Education: A Systematic Mapping of the Literature

Indianara Amâncio
Universidade do Estado de Santa
Catarina
ORCID:[0000-0002-7547-9840](https://orcid.org/0000-0002-7547-9840)
indianara.amancio@gmail.com

Avanilde Kemczinski
Universidade do Estado de Santa
Catarina
ORCID:[0000-0001-7671-5457](https://orcid.org/0000-0001-7671-5457)
avanilde.kemczinski@udesc.br

Marlow R. B. Dickel
Universidade do Estado de Santa
Catarina
ORCID:[0000-0002-1365-6442](https://orcid.org/0000-0002-1365-6442)
profmarlow@gmail.com

Resumo

A autoavaliação discente é uma estratégia de avaliação que proporciona benefícios educacionais aos estudantes, como a capacidade de avaliar seu próprio trabalho e a autorregulação de seu aprendizado. O objetivo desta pesquisa é apresentar como a autoavaliação discente vem sendo aplicada no ensino superior por meio de um Mapeamento Sistemático da Literatura (MSL) sobre o tema. Os quarenta e seis trabalhos relevantes mostraram que a autoavaliação normalmente é aplicada junto a outras estratégias, como avaliação em pares e do professor. Ganhos educacionais como melhora no desempenho acadêmico, autocrítica, reflexão e responsabilidade foram observados pela prática da autoavaliação. Ainda, dos 53% dos trabalhos que mencionaram o uso de ferramentas computacionais no emprego da autoavaliação, observou-se que foram necessárias duas ferramentas, uma para aplicação da autoavaliação e outra para análise de dados. Já em relação à acurácia, não foi percebido um consenso por parte dos autores, sendo que 26% mencionam utilizar o coeficiente de correlação de Pearson, que não atende a todos os cenários possíveis na comparação de estratégias como a autoavaliação e avaliação do professor. Um coeficiente que permite mensurar a acurácia da autoavaliação discente é sugerido nesta pesquisa, haja vista que a dificuldade em mensurar acurácia foi uma lacuna observada.

Palavras-Chave: autoavaliação discente; coeficiente de acurácia; ensino superior; mapeamento sistemático da literatura.

Abstract

Student self-assessment is an assessment strategy that provides educational benefits to students, such as the ability to evaluate their own work and self-regulate their learning. The purpose of this research is to present how student self-assessment has been applied in higher education through a Systematic Mapping of the Literature (SML) on the topic. The forty-six relevant papers showed that self-assessment is usually applied along with other strategies such as peer and teacher assessment. Educational gains such as improved academic performance, self-criticism, reflection, and accountability were observed by practicing self-assessment. Furthermore, of the 53% of the studies that mentioned the use of computer tools in the use of self-assessment, it was observed that two tools were necessary, one for the application of self-assessment and the other for data analysis. In relation to accuracy, no consensus was found among the authors, with 26% mentioning the use of Pearson's correlation coefficient, which does not meet all possible scenarios when comparing strategies such as self-assessment and teacher evaluation. A coefficient that allows measuring the accuracy of student self-assessment is suggested in this research, since the difficulty in measuring accuracy was a gap observed.

Keywords: student self-assessment; coefficient of accuracy; higher education; systematic mapping of the literature.

1 Introdução

O envolvimento dos estudantes na identificação de padrões ou critérios para aplicar em seu próprio trabalho e fazer julgamentos sobre ele é uma definição de Boud (1995) sobre a autoavaliação discente. A capacidade de se autoavaliar traz benefícios tanto na academia quanto no âmbito profissional do indivíduo. Isso acontece quando a autoavaliação é colocada como um aspecto crítico para a avaliação da aprendizagem e ao longo da vida (Tan, 2008). Na academia, a autoavaliação ajuda a promover a autoeficácia. Vieira (2013, p.28) reconhece que “o aluno que se autoavalia revela um grande envolvimento pessoal no processo de avaliação, gerindo o seu desempenho e refletindo sobre ele, tornando-se assim regulador da sua própria aprendizagem”.

Os MSLs sobre autoavaliação discente encontrados previamente possuem diferentes focos e foram realizados em períodos distantes de tempo. Sendo assim, a condução desta pesquisa fez-se necessária para que se pudesse ter uma visão atualizada sobre como essa estratégia de avaliação vem sendo aplicada no ensino superior. Foram observados aspectos como: quais ferramentas vêm sendo utilizadas em sua aplicação (rubricas, questionários, etc); quais outras estratégias de avaliação estão sendo usadas em conjunto com a autoavaliação; quais *softwares* vêm sendo utilizados para automatizar a aplicação da autoavaliação; quais ganhos educacionais a autoavaliação vem proporcionando e como vem sendo medida a acurácia da autoavaliação. Para o atingimento desse objetivo, um MSL foi conduzido entre os meses de abril e maio do ano de 2021.

Uma lacuna em relação à existência de um *software* que reúna os requisitos de aplicação da autoavaliação e visualização dos dados foi observada, além de uma não conformidade em relação à mensuração da acurácia da autoavaliação. Alguns autores inferiram dificuldades em mensurar a acurácia, outros não discutiram sobre esse aspecto por sentirem a necessidade de maior entendimento e prática por parte dos estudantes. Em observância à essa lacuna, esta pesquisa apresenta uma fórmula para o cálculo de um coeficiente de acurácia de autoavaliação, levando em consideração as pontuações de uma rubrica de autoavaliação em comparação com a rubrica de avaliação do professor.

O trabalho está organizado da seguinte forma: na fundamentação teórica, os conceitos importantes que permeiam o tema são definidos; na seção de metodologia, o protocolo do MSL é apresentado, bem como as questões secundárias são respondidas, embasando a resposta à questão primária de pesquisa: “como vem ocorrendo o processo da aplicação da autoavaliação discente no ensino superior?”; já na seção “coeficiente de acurácia de autoavaliação”, o cálculo do coeficiente proposto e a explicação de como ele deve ser aplicado são apresentados. Por fim, na seção de conclusão, as reflexões acerca da pesquisa são apresentadas, além de mencionadas ameaças à validade deste trabalho e direcionamentos à comunidade de informática na educação.

2 Fundamentação teórica

Segundo Monteiro, Leitão e Barreto (2021) a avaliação é “tarefa corriqueira no âmbito da educação”, mas nem sempre auxilia efetivamente o estudante a entender como está se desenvolvendo. Na pesquisa de Oliveira e Santos (2005), uma análise sobre a avaliação de aprendizagem nas universidades foi realizada. As autoras colocam que não há uma coerência entre os autores no que se refere à avaliação, parecendo haver um descontentamento em relação à passividade do estudante nesse processo. Acrescentam ainda que seria interessante que se repensasse o discurso sobre essa passividade, buscando pontos de convergência sobre o conceito e melhores formas de avaliação.

2.1 Autoavaliação discente

Uma forma de avaliação que busca a ação ao invés da passividade do estudante, é a autoavaliação, principalmente no contexto das metodologias ativas onde se faz ainda mais necessária essa mudança de postura. Ao encontro dessa afirmação vêm as ideias de Bacich e Moran (2018, p. 86), quando colocam que “ao pensarmos em mudanças no contexto educacional, não podemos nos esquecer de rever os modelos de avaliação. De um modo geral, sempre que se faz menção aos processos avaliativos, esses são considerados unilateralmente, ou seja, apenas do professor em relação aos alunos, quando deveriam ser concebidos em via dupla, além de abarcar um espaço para a autoavaliação, que deve ser realizada pelo aluno e também pelo professor”. Em consonância com essa reflexão, Vieira e Moreira (2003, p. 33) inferem que “considerar a autonomia como meta de aprendizagem implica salientar a importância da dimensão reflexiva dessa aprendizagem e a competência processual do aluno”.

Promover a autorregulação no processo de ensino-aprendizagem é uma tarefa importante, pois como colocam Montalvo e Torres (2004, p.04) “se há algo que caracteriza esses alunos é que se sentem agentes de sua conduta, acreditam que a aprendizagem é um processo proativo, estão motivados e usam as estratégias que lhes permitem alcançar os resultados acadêmicos desejados”. Vieira (2013) reconhece que “o aluno que se autoavalia revela um grande envolvimento pessoal no processo de avaliação, gerindo o seu desempenho e refletindo sobre ele, tornando-se assim regulador da sua própria aprendizagem”. A autoavaliação pode desempenhar um papel facilitador na promoção da autorregulação da aprendizagem.

Para que a autoavaliação atinja os objetivos educacionais que pode promover, é necessária a observação de vários aspectos, sendo dois deles, segundo Boud (1995): que os estudantes e a equipe como um todo estejam cientes dos critérios e julgamentos e que façam parte da etapa de definição desses critérios. A pesquisa em autoavaliação está predominantemente focada na eficácia da promoção do desempenho acadêmico, mas pouco focada na validade e confiabilidade dessa estratégia de avaliação (Brown, Andrade, Chen; 2015; Han, 2018).

2.2 Acurácia da autoavaliação discente

Boud, Lawson e Thompson (2013) inferem que a acurácia da autoavaliação está pautada nos alunos exercerem essa estratégia constantemente e relacionarem suas pontuações com as pontuações dos professores e colegas, buscando entender onde podem melhorar. Han (2018) concorda com os autores e faz menção a outras pesquisas (Quadro 1) que já partem do pressuposto que a acurácia da autoavaliação está diretamente relacionada com o quão ela se aproxima da avaliação do professor.

Quadro 1: Pesquisas que tratam a acurácia da autoavaliação.

Autor(es) e ano	Título
Blanche e Merino (1989)	<i>Self-Assessment of Foreign-Language Skills: Implications for Teachers and Researchers</i>
Boud e Falchikov (1989)	<i>Student Self-Assessment in Higher Education: A Meta-Analysis</i>
Lew, Alwis e Schmidt (2010)	<i>Accuracy of students' self-assessment and their beliefs about its utility</i>

Andrade (2019) em sua pesquisa defende o aspecto formativo como o mais adequado, uma vez que entende a autoavaliação diretamente relacionada ao *feedback*. Além disso, a autora menciona que, como qualquer habilidade que é apoiada e aprendida, a acurácia da autoavaliação melhora ao longo do tempo. Hale (2015) acrescenta que estudantes e professores devem usar a mesma ferramenta de avaliação, evitando a impressão de que somente a ferramenta utilizada pelo professor é válida para fins somativos.

2.3 Correlação de Pearson

Partindo-se do princípio de que a acurácia da autoavaliação discente está relacionada ao quão seus resultados são parecidos com a avaliação do professor, é necessário que essa medição seja feita por um método eficaz. Considerar apenas a média final das avaliações empregadas pode não refletir a proximidade ou distância das estratégias. Sendo assim, a correlação de Pearson é uma alternativa que pode auxiliar na análise dessa proximidade. Para Garson (2009, p.1), “é uma medida de associação bivariada (força) do grau de relacionamento entre duas variáveis”. A fórmula da correlação de Pearson é dada por:

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{n\sum x^2 - (\sum x)^2} \sqrt{n\sum y^2 - (\sum y)^2}} \quad (1)$$

Larson e Farber (2015) colocam que o coeficiente tem como resultado um valor de -1 a +1, sendo que se “x” e “y”, as duas variáveis a serem correlacionadas, têm uma correlação linear positiva forte, “r”, que corresponde ao coeficiente de correlação, está perto de 1. Porém, se “x” e “y” têm uma correlação linear negativa forte, “r” está perto de -1.

3 Metodologia da pesquisa

Para Petersen, Vakkalanka e Kuzniarz (2015) o MSL é uma técnica que visa proporcionar uma visão geral sobre uma determinada área de pesquisa. Kitchenham e Charters (2007) o colocam como uma ampla revisão de trabalhos primários em uma área para a identificação das evidências disponíveis sobre o tema. A definição do protocolo de pesquisa é uma etapa importante desse tipo de trabalho, como colocam Dermeval, Coelho e Bittencourt (2019), também contemplando as etapas mostradas na Figura 1.



Figura 1: Etapas do MSL.

Um MSL é considerado por Dermeval, Coelho e Bittencourt (2019) como um dos tipos de trabalho secundário. Os autores definem como trabalho secundário aquele que “revisa os estudos primários referentes a uma questão de pesquisa específica com o objetivo de integrar/sintetizar as evidências relacionadas à questão de pesquisa”.

As próximas seções descrevem como cada uma das etapas do protocolo de pesquisa do mapeamento foi desenvolvida.

3.1 Identificação da necessidade do mapeamento

Para Derneval, Coelho e Bittencourt (2019) a primeira falha de alguns pesquisadores de informática na educação é não verificar de forma adequada se o mapeamento que pretendem conduzir é realmente necessário. Isso significa buscar por pesquisas já realizadas a fim de aproveitar os conhecimentos já discutidos por outros autores e concentrar esforços nas lacunas observadas. Sendo assim, foram pesquisados no Google Scholar trabalhos em inglês e português que executaram algum tipo de Revisão Sistemática da Literatura (RSL) ou MSL, ambos caracterizados como trabalhos secundários. Os trabalhos dispostos no Quadro 2 foram encontrados e na sequência, as discussões e resultados obtidos por cada um são apresentados.

Quadro 2: Trabalhos secundários sobre autoavaliação discente.

Autores e ano	Título
Dochy, F.; Segers, M.; Sluijsmans, D. (1999)	<i>The use of self-peer and co-assessment in higher education: A review</i>
França, Rozelma Soares De; Tedesco, Patrícia Cabral De Azevedo Restelli (2015)	Caracterizando a pesquisa sobre autoavaliação na aprendizagem de programação para iniciantes
Panadero, Ernesto; Brown, Gavin T. L.; Strijbos, Jan-Willem (2015)	<i>The Future of Student Self-Assessment: a Review of Known Unknowns and Potential Directions</i>

Na pesquisa de Dochy, Segers e Sluijsmans (1999), foi realizada uma RSL, baseando-se na análise de sessenta e três estudos sobre o uso da avaliação em pares, autoavaliação e coavaliação no ensino superior. Três questões nortearam a pesquisa: a) quais são as principais contribuições da pesquisa sobre novas formas de avaliação como a autoavaliação, avaliação em pares e coavaliação? b) de que forma os resultados podem ser reunidos? c) quais diretrizes podem ser dadas aos profissionais da educação? Os resultados mostraram que existem evidências de que a participação dos estudantes no processo de avaliação é algo importante. Sobre a autoavaliação, Dochy, Segers e Sluijsmans (1999) mencionam que a literatura se pauta em seis aspectos: a influência das diferentes habilidades do estudante na acurácia da autoavaliação; o efeito do tempo; a acurácia da autoavaliação em relação à avaliação do professor; o efeito da autoavaliação; métodos de autoavaliação e o conteúdo da autoavaliação.

Sobre a influência das diferentes habilidades, Boud e Falchikov (1989) *apud* Dochy, Segers e Sluijsmans (1999), analisaram estudos entre 1932 e 1988 nos quais investigaram as comparações da autoavaliação dos estudantes. Os autores reportaram que os bons estudantes tendem a se superestimar e os estudantes com maiores dificuldades tendem a se subestimar. Sobre o efeito do tempo, Griffiee (1995) *apud* Dochy, Segers e Sluijsmans (1999) investigaram se existe diferença na aplicação da autoavaliação discente no primeiro, segundo ou terceiro ano na universidade. A conclusão geral do autor é que não há diferença entre os anos e, além disso, a interpretação dos estudantes depende da forma como a autoavaliação é inserida no contexto e das experiências educacionais. Quanto à acurácia da autoavaliação em relação à avaliação do professor, Zoller & Ben-Chaim (1997) *apud* Dochy, Segers e Sluijsmans (1999) verificaram se os alunos eram capazes de se autoavaliar de forma acurada. Os autores concluíram que os estudantes em geral foram precisos, mas que os estudantes menos motivados foram os menos precisos dentro dos critérios estabelecidos.

Já em França e Tedesco (2015), um MSL foi conduzido para identificar o uso da autoavaliação no ensino de programação para iniciantes no ensino superior. Os onze trabalhos resultantes foram classificados como: experiência didática, ferramenta e estudo teórico. Em relação às ferramentas utilizadas, em quatro trabalhos um questionário ou formulário foi utilizado e um trabalho utilizou um *quiz*. Em relação aos *softwares* utilizados, nos onze trabalhos resultantes do mapeamento de França e Tedesco (2015), no trabalho de Kay et al. (2007) *apud* França e Tedesco (2015), foi usado o Reflect. Nessa ferramenta, o estudante avalia soluções propostas a

partir de critérios estabelecidos pelo professor. A avaliação do estudante é comparada à avaliação do professor e a discrepância mostra o quão longe o estudante está do desejado. No entanto, essa ferramenta apenas calcula o total da diferença de pontuação entre professor e estudante.

Outro *software* é mencionado no trabalho de Schorsch (1995) *apud* França e Tedesco (2015), que fornece *feedbacks* aos estudantes sobre erros de código em relação à sintaxe e lógica. França e Tedesco (2015, p. 555) explicam que os *feedbacks* do *software* “informam o que está errado, incluindo uma amostra de código correto como exemplo para o aprendiz”. Os resultados obtidos demonstram que nos últimos anos houve um aumento do interesse pela temática, que a maioria das experiências se concentram no ensino superior e que os alunos foram capazes de se autoavaliar com precisão.

Na pesquisa de Panadero, Brown e Strijbos (2015), uma revisão sobre autoavaliação e direções futuras do emprego dessa estratégia foram discutidas, caracterizando as tipologias de autoavaliação (formatos), acurácia, expectativas e seus efeitos. Como resultados, os autores destacam que a acurácia da autoavaliação é relacionada positivamente com o bom desempenho do estudante e que melhora com o passar do tempo. Em relação às tipologias, mencionam que são necessários mais estudos para caracterizar a facilidade de aplicação e os ganhos educacionais de cada uma. Sobre as contribuições da autoavaliação, é destacada a autorregulação do aprendizado.

Existiu a necessidade de se realizar um MSL atualizado, pois a pesquisa de Dochy, Segers e Sluijsmans (1999) é bastante antiga e não analisa *softwares* que foram utilizados para o emprego da autoavaliação discente. Já na pesquisa de França e Tedesco (2015), existe a investigação sobre os *softwares* utilizados, mas a pesquisa limita-se a analisar a autoavaliação no contexto do ensino de programação. Por fim, a pesquisa de Panadero, Brown e Strijbos (2015) enfatiza as tipologias de autoavaliação, mas também não investiga os *softwares* utilizados. Sendo assim, concluiu-se que seria necessário realizar essa pesquisa para se obter o panorama atualizado de como a autoavaliação vem sendo aplicada no ensino superior.

3.2 Formulação de uma questão de pesquisa

Para nortear os estudos, foi definida uma Questão Principal de Pesquisa (QPP) e cinco Questões Secundárias de Pesquisa (QSP), sendo elas:

- **QPP:** Como vem ocorrendo o processo da aplicação da autoavaliação discente no ensino superior?
 - O objetivo da QPP é discutir de forma geral, considerando as respostas às questões secundárias, como a autoavaliação discente vem sendo aplicada no ensino superior, enfatizando a observação e comparação dos *softwares* que têm apoiado o processo.
- **QSP01:** Quais ferramentas e critérios de avaliação vêm sendo aplicados nas autoavaliações?
 - O objetivo dessa QSP é identificar de que maneira a autoavaliação vem sendo aplicada no que diz respeito aos critérios utilizados e as ferramentas (ex.: rubrica, questionário, *checklist*).
- **QSP02:** Quais outras estratégias de avaliação vêm sendo aplicadas em conjunto com a autoavaliação?
 - O objetivo dessa QSP é analisar que outras estratégias de avaliação (ex.: avaliação em pares, avaliação do professor) vêm sendo aplicados junto à autoavaliação, possibilitando entender o papel específico da autoavaliação no processo.

- **QSP03:** Quais *softwares* vêm sendo utilizados para dar suporte ao uso da autoavaliação?
 - O objetivo dessa QSP é identificar, comparar e caracterizar os *softwares* que vêm sendo utilizados para apoiar a aplicação e integração da autoavaliação no processo de ensino-aprendizagem na educação superior.
- **QSP04:** Quais são as contribuições e os impactos que a autoavaliação vem proporcionando no processo de ensino-aprendizagem no ensino superior?
 - O objetivo dessa QSP é analisar os ganhos educacionais que a autoavaliação discente propicia ao estudante, aos pares, ao professor e ao processo de ensino-aprendizagem como um todo.
- **QSP05:** Como foi medida a acurácia da autoavaliação discente?
 - O objetivo dessa QSP é entender como a acurácia da autoavaliação é entendida pelos autores e como foi medida.

Na etapa de extração de dados, a análise dos artigos foi realizada considerando as questões secundárias de pesquisa para que se pudesse responder à questão primária de pesquisa definida.

3.3 Busca e seleção dos estudos primários

Os mecanismos de busca utilizados para a pesquisa foram baseados no trabalho de Buchinger, Cavalcanti e Hounsell (2014), onde uma análise quantitativa dos recursos disponíveis em 40 mecanismos foi realizada. De acordo com os autores, os mecanismos melhor avaliados foram: Web of Knowledge, Engineering Village, Scopus SciVerse, IEEE Xplore, ACM Digital Library, Science Direct e Springer Link, e por isso foram os utilizados no MSL. A frase de busca foi composta pela utilização dos termos *self-assessment* e *higher education*, dadas as devidas adaptações nos mecanismos de busca e sinônimos, tendo como resultado: (“*self-assessment*” OR “*self-evaluation*”) AND (“*higher education*” OR “*university*” OR “*college*”). A pesquisa foi realizada pelo título e resumo do documento e o intervalo de anos considerado e aplicado diretamente em cada mecanismo foi de 2008 a 2020.

3.4 Seleção de qualidade

Os critérios de inclusão e exclusão são definidos para que se possam extrair os trabalhos relevantes dentre os resultantes das pesquisas. O Quadro 3 exhibe os critérios de inclusão e exclusão definidos para este mapeamento.

Quadro 3: Critérios de inclusão e exclusão.

Tipo	Id	Descrição
Inclusão	CI1	Artigos escritos em inglês
	CI2	Artigos primários
	CI3	Artigos completos (<i>full paper</i>)
	CI4	Artigos publicados em revistas ou conferências
Exclusão	CE1	Artigos que não aplicam autoavaliação discente para fins educacionais
	CE2	Artigos que não se referem ao ensino superior
	CE3	Artigos não disponíveis para <i>download</i> ou acesso na íntegra
	CE4	Artigos duplicados

Quando pela leitura do título do documento e resumo o trabalho era considerado como possivelmente relevante, era marcado como “para revisão”, onde uma leitura mais ampla dos capítulos de metodologia, resultados e conclusão era realizada. Nessa leitura, os critérios de inclusão e exclusão eram novamente observados a fim de garantir que o trabalho realmente era relevante. Em relação ao critério de exclusão CE1, ele foi criado para excluir trabalhos que aplicavam a autoavaliação discente para fins sociais e não educacionais, ou aplicavam a autoavaliação da instituição.

Considerando as pesquisas realizadas em todos os mecanismos, inicialmente foram encontrados oitocentos e vinte e três trabalhos, que foram organizados em uma planilha do Microsoft Excel. Com os critérios de inclusão e exclusão aplicados, foram selecionados quarenta e seis trabalhos relevantes. A Tabela 1 exibe a quantidade de trabalhos retornados e relevantes em cada mecanismo.

Tabela 1: Trabalhos retornados e relevantes por mecanismo de busca.

Mecanismo	Retornados	Relevantes
ACM DL	11	5
Springer	21	7
Scopus	21	1
Engineering Village	37	4
Science Direct	116	9
Web of Science	196	12
IEEE Xplore	421	8
Total	823	46

No Apêndice A estão listados os quarenta e seis trabalhos relevantes. Visando novas perspectivas sobre os trabalhos, foram gerados dois gráficos: publicações por ano e por continente de publicação. A seguir, é mostrada a Figura 2, considerando os anos de publicação.

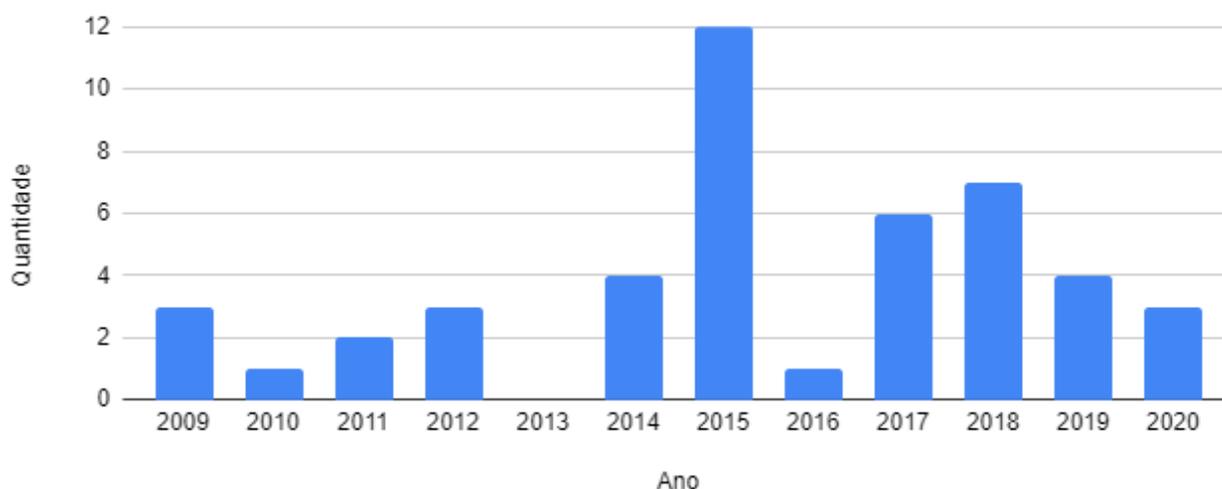


Figura 2: Trabalhos relevantes por ano.

Além de se fazer evidente o ano com mais publicações relevantes (2015), pode-se notar que a maior parte dos resultados aparece na metade mais recente dos anos de pesquisa, e que, após o hiato de 2013, apenas um ano teve menos de três publicações (2016). Isso demonstra a relativa proximidade temporal dos trabalhos selecionados. Na Figura 3, é apresentado um comparativo entre os continentes onde esses artigos foram publicados.

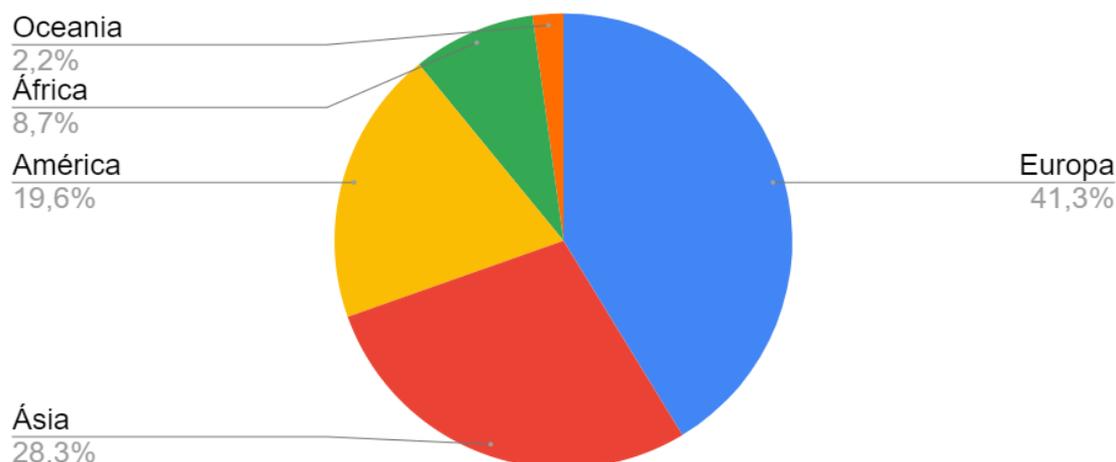


Figura 3: Trabalhos relevantes por continente.

Nota-se que a maior parte das publicações vêm da Europa, com dezenove publicações, das quais nove foram realizadas na Espanha, país com mais publicações selecionadas. Em seguida, aparece o continente asiático, com quatorze publicações distribuídas em sete países, sendo a China destaque com quatro. O terceiro continente é o americano, com oito publicações nos Estados Unidos e uma no Canadá, ou seja, concentrando as publicações ao norte do continente, o que possibilita destacar que não há uma proximidade geográfica deste trabalho com os determinados como relevantes. África, com quatro, e Oceania, com um artigo na Austrália, completam a lista.

3.5 Extração de dados, sumarização e síntese dos resultados

Nesta seção, as respostas às questões de pesquisa e outros dados observados e extraídos dos trabalhos relevantes são discutidos. A primeira **QSP01** definida preocupava-se em saber quais ferramentas e critérios de avaliação vêm sendo aplicados nas autoavaliações.

3.5.1 QSP01- Ferramentas e critérios de autoavaliação

A Tabela 2 contabiliza entre os quarenta e seis trabalhos relevantes a quantidade de trabalhos por ferramenta de autoavaliação utilizada.

Tabela 2: Ferramentas de autoavaliação.

Ferramenta	Quantidade de trabalhos
Questionário avaliando competências	28
Questionário com questões técnicas	11
<i>Feedback</i>	03
Rubrica	03
Não mencionada	01

Mais da metade dos trabalhos (61%) utilizou como ferramenta de autoavaliação discente o questionário contendo questões que avaliavam competências. Na pesquisa de Perera, Mohamadou e Kaur (2010), perguntas como “o que eu fiz bem durante a entrevista para conseguir um bom relacionamento com o paciente?” e “eu fiz uso de linguagem apropriada?” foram usadas no questionário de autoavaliação. Já na pesquisa de Pop e Khampirat (2019), os estudantes responderam questões de vinte aspectos diferentes, entre eles: habilidades de comunicação, habilidades de escrita, habilidades interpessoais, habilidades de tomada de decisão e resolução de problemas. De forma semelhante, em Heukelman (2017) os estudantes autoavaliaram suas competências em diferentes aspectos como: competências gerais da internet, mídias sociais, criação de documentos, *e-commerce* e armazenamento em nuvem.

Em 24% dos trabalhos relevantes, exercícios ou questionários técnicos foram utilizados como ferramenta de autoavaliação. Na pesquisa de Mendonza et al. (2016), os estudantes respondiam questões técnicas sobre programação em um aplicativo desenvolvido pela universidade e já recebiam o *feedback* instantâneo sobre a resposta estar correta ou não. Dessa forma, autoavaliaram seus conhecimentos em programação. Em Hwang et al. (2015), os estudantes também eram instigados a se autoavaliarem respondendo questões de programação. Perguntas como: “você poderia mostrar forças ou fraquezas do *loop* aninhado e mostrar a diferença ou similaridade com outra estrutura?” foram usadas no questionário de autoavaliação. O questionário utilizado pedia exemplos de códigos de programação em diversas questões. Em Matsuno (2009), os estudantes respondiam em escala Likert de seis pontos o quão foram bem em uma redação, avaliando aspectos técnicos como: ortografia, pontuação, sequência lógica, relevância do tema, entre outros.

A rubrica foi uma ferramenta de autoavaliação utilizada em 7% dos trabalhos relevantes. Em Malhberg (2015), no curso de fotografia, o professor elaborou uma rubrica com questões que envolviam qualidade, estética e fluxo de trabalho. Já na pesquisa de González-Betancor, Cruz e Verano-Tacoronte (2017), foi criada uma rubrica para que os estudantes autoavaliassem suas apresentações orais. Da mesma forma, em três trabalhos, que também representam 7% dos trabalhos relevantes, o *feedback* foi utilizado como ferramenta de autoavaliação. Em Hulsman e Vloodt (2015), os estudantes postaram um vídeo com uma consulta simulada com um paciente. Como autoavaliação, deveriam postar comentários positivos e negativos sobre seus vídeos. Já em Pereira et al. (2014), além de uma autoavaliação de seus vídeos considerando clareza, precisão, linguagem e estrutura visual entre outros aspectos, os estudantes também foram instigados a deixar um comentário em suas autoavaliações e nas autoavaliações dos colegas.

No contexto de como foi solicitado que respondessem a essas ferramentas de autoavaliação, houve uma predominância por respostas em escalas, como é apresentado na Tabela 3:

Tabela 3: Formas de resposta aos critérios de avaliação.

Tipo de resposta	Quantidade de trabalhos
Escala Likert	16
Não mencionado	9
Escala numérica	7
Outro tipo de escala	6
Resposta aberta	4
<i>Feedback</i>	3
Resposta técnica	1

Escalas foram utilizadas em 63% dos trabalhos relevantes, sendo a escala Likert utilizada em dezesseis, correspondendo a 35% do total. Segundo Silva Júnior e Costa (2014, p.4-5), que em sua pesquisa fazem uma análise comparativa entre escalas, a escala Likert é o modelo mais utilizado e debatido entre os pesquisadores, e foi desenvolvido por Rensis Likert em 1932 para mensurar atitudes no contexto das ciências comportamentais. A escala de verificação de Likert consiste em tomar um construto e desenvolver um conjunto de afirmações relacionadas à sua definição, para as quais os respondentes emitirão seu grau de concordância.

No trabalho de Hosein e Harle (2018), os estudantes preencheram por meio de uma escala Likert de cinco pontos, variando de “1 - não muito confiante” até “5 - muito confiante”, seu grau de confiança nas perguntas estabelecidas. Já em Toro, Traver e Amengual (2014), os estudantes responderam em uma escala Likert, de “1 - discordo completamente” até “5 - concordo completamente”, a afirmações sobre estudo e desenvolvimento, honestidade e trabalho em equipe e competências técnicas. Por exemplo, a afirmação “o trabalho entregue foi totalmente desenvolvido pelos membros da equipe” foi uma das apresentadas. Também em escala Likert de cinco pontos, os estudantes na pesquisa de Lewis, Titterton e Clancy (2011) responderam o quão estavam seguros em relação à conhecimentos técnicos de programação como ponteiros em

linguagem C, considerando “1 - não entendi o tópico ao todo”, até “5 - estou completamente confortável”. Ainda, em nove trabalhos, apesar de se ter aplicado a autoavaliação discente e coletado as respostas, não fica clara a escala ou forma de resposta aos critérios definidos.

A seção seguinte discute a **QSP02**, onde o emprego de outras estratégias de avaliação junto à autoavaliação foi observado.

3.5.2 QSP02- Outras estratégias de avaliação

Nessa questão secundária de pesquisa, foi observada a ocorrência da aplicação de outras estratégias de avaliação, como avaliação em pares e do professor, junto com a autoavaliação. A Tabela 4 contabiliza o número de trabalhos e as estratégias de avaliação mencionadas:

Tabela 4: Estratégias de avaliação.

Estratégias	Quantidade de trabalhos
Autoavaliação somente	14
Autoavaliação e outras atividades	11
Autoavaliação e avaliação em pares	08
Autoavaliação, avaliação em pares e avaliação do professor	07
Autoavaliação e avaliação do professor	06

Nos onze trabalhos (24%) que utilizaram outras atividades além da autoavaliação, pode-se citar como exemplo a pesquisa de Mendoza et al. (2014) que menciona ter utilizado uma avaliação do projeto final do curso, onde foi observado que os estudantes que praticaram a autoavaliação estiveram melhor preparados para tal exame. Já na pesquisa de Hosein e Harle (2018), além da autoavaliação, foi aplicado um teste de matemática, onde foi observado que os estudantes com mais dificuldades acabaram sendo mais precisos em suas autoavaliações.

Se somadas as quantidades de trabalhos que utilizaram autoavaliação e avaliação em pares, autoavaliação e avaliação do professor ou as três estratégias em conjunto, tem-se um total de 46%. Foram realizadas comparações entre os tipos de avaliação utilizados, em relação aos trabalhos que envolveram a avaliação do professor e avaliação em pares. Em Hulsman e Vloodt (2015), as autoavaliações continham mais texto e eram mais específicas do que as avaliações em pares. Já em Han (2018), a avaliação em pares e avaliação do professor em geral tiveram mais correlação do que a autoavaliação e a avaliação em pares. O mesmo aconteceu em Aryadoust (2015), onde também as autoavaliações tiveram pontuações mais baixas.

Em Pereira et al. (2014), os estudantes se interessaram mais pela avaliação em pares, mas os que fizeram a autoavaliação procuraram ser responsáveis e críticos, onde muitas vezes as pontuações eram inferiores às avaliações do professor e em pares.

3.5.3 QSP03- Softwares utilizados

Os *softwares* utilizados para aplicação da autoavaliação e análise de dados foram um aspecto observado nos trabalhos relevantes. É importante ressaltar que no Quadro 4 são mencionados apenas os artigos que falaram sobre a utilização de alguma ferramenta computacional e qual o objetivo da ferramenta: aplicação da autoavaliação ou análise de dados, compreendendo assim vinte e dois (48%) dos quarenta e seis (100%) artigos.

Quadro 4: Softwares utilizados.

Autor(es) e ano	Software	Autoavaliação	Análise de dados
Diogo, Barbosa e Ferreira (2015)	Google Forms	X	-
	Excel	-	X
	SPSS	-	X
Dlab, Katic, Candrljic (2015)	Moodle	X	-
Duran et al. (2019)	Não especificado	X	-
Guijarro e Bengoechea (2017)	Não especificado	X	-
Hale (2015)	Não especificado	X	-
	HyperSearch 3.0	-	X
Heukelman (2017)	IATA	-	X
Hulsman e Vloodt (2014)	VFR	X	-
	Excel	-	X
Hwang et al. (2015)	VPen	X	-
Ibabe e Jauregizar (2010)	Hot Potatoes	X	-
Lanthony et al. (2018)	Moodle	X	-
Lladó et al. (2020)	SPSS	-	X
Matsuno (2009)	FACETS	-	X
Mendoza et al. (2015)	iPractice app	X	-
Meng, Cui e Wang (2015)	Não especificado	X	-
	PASW Estatística	-	X
Othman e Abdullah (2018)	Google Forms	X	-
	Excel	-	X
Pereira et al. (2014)	Babelium (Moodle)	X	-
Pop e Khampirat (2019)	E-mail	X	-
	SPSS	-	X
Sánchez, Soldado e López (2014)	Moodle	X	-
Ssegawa e Kasule (2017)	Não especificado	-	X
Walser (2009)	Blackboard	X	-
Weisi e Karimi (2012)	Não especificado	-	X
Yan et al. (2011)	Não especificado	X	-

Yan et al. (2011), Hale (2015), Meng, Cui e Wang (2015), Guijarro e Bengoechea (2017) e Duran et al. (2019) não especificam qual *software* utilizaram para a aplicação da autoavaliação, mas sugerem ter realizado por um meio computacional. No entanto, Meng, Cui e Wang (2015) utilizaram o *software* PASW Estatística para a análise de dados, bem como, Hale (2015) utilizou o HyperSearch 3.0 para o mesmo fim.

Já em Sánchez, Soldado e López (2014), Lanthony et al. (2018) e Dlab, Katic, Candrljic (2015), o Moodle é citado como o *software* usado para a aplicação da autoavaliação. Já na pesquisa de Pereira et al. (2014) foi citado o *software* Babelium, que é um *plugin* para o Moodle, onde os estudantes postaram seus vídeos.

Na pesquisa de Diogo, Barbosa e Ferreira (2015) e Othman e Abdullah (2018) o Google Forms foi utilizado para aplicar o questionário de autoavaliação com os estudantes. Em ambas as pesquisas, foi utilizada também a ferramenta Microsoft Excel para a análise de dados e em Diogo, Barbosa e Ferreira (2015) também o *software* estatístico IBM *Statistical Product and Service Solutions* (SPSS). Em Hulsman e Vloodt (2014), a ferramenta *Video Fragment Rating* (VFR), que se trata de uma plataforma para postar trechos de vídeos clínicos por estudantes e professores, foi utilizada. Os autores também mencionam ter utilizado o Microsoft Excel para análise de dados.

Em Hwang et al. (2015), a ferramenta *Virtual Pen* (VPen) foi utilizada para dar suporte à aplicação da autoavaliação, onde os estudantes compartilhavam suas anotações sobre os conteúdos. Na pesquisa de Ibabe e Jauregizar (2010) a ferramenta Hot Potatoes foi utilizada. Trata-se de um pacote que inclui seis aplicativos que permitem a criação de diversos tipos de exercícios de múltipla escolha, frases curtas, palavras cruzadas, entre outros. Pop e Khampirat

(2019) mencionaram utilizar o e-mail para coleta de respostas das autoavaliações e o *software* estatístico IBM SPSS para análise de dados.

Em Matsuno (2009), o aplicativo proprietário iPractice foi desenvolvido para coletar as respostas às questões técnicas de programação e fornecer o *feedback* imediato de acerto ou erro ao estudante. Na pesquisa de Walser (2009), a autoavaliação foi disponibilizada aos estudantes por meio de uma ferramenta da empresa Blackboard. No entanto, a autora menciona ter feito análises estatísticas sobre as respostas, mas não divulga se conseguiu fazer tais análises pelo próprio Blackboard ou se necessitou de outro *software*.

Weisi e Karimi (2012) e Ssegawa e Kasule (2017) sugerem ter utilizado ferramentas para análise de dados, mas não explicitam quais. Em contrapartida, Heukelman (2017), Matsuno (2009) e Lladó et al. (2020) mencionam ter utilizado para análise de dados respectivamente os *softwares*: *Item And Test Analysis* (IATA), FACETS e IBM SPSS.

3.5.4 QSP04- Contribuições da autoavaliação

As contribuições observadas pelos autores após o emprego da autoavaliação em seus experimentos foram variadas. Em alguns trabalhos, houve uma correlação entre as contribuições ao processo de ensino-aprendizagem, como é mostrado na Tabela 5.

Tabela 5: Contribuições da autoavaliação.

Contribuições	Quantidade de trabalhos
Melhor desempenho acadêmico	10
Reflexão acerca de seu trabalho	07
Autorregulação do aprendizado	05
Autoconfiança	04
Autocrítica	04
Maior motivação	04
Responsabilidade	03
Monitoramento do processo de ensino-aprendizagem	03
Conscientização	02

Foram mencionadas a melhor assimilação da teoria e prática pelos estudantes em Dlab, Katic, Candrljic (2015), melhor controle e gerenciamento do trabalho em Ezzahra, Islam e Radid (2015), maior participação nas aulas em Sánchez, Soldado e López (2014), maior persistência de estudantes em Mahlberg (2015), maior perseverança na realização das atividades acadêmicas em Lewis, Titterton e Clancy (2011) e melhor desempenho dos estudantes segundo a percepção dos professores em Othman e Abdullah (2018).

É importante colocar que alguns autores foram inconclusivos ou não dissertaram sobre o impacto positivo da autoavaliação. Como exemplo, na pesquisa de Mbonigaba e Oumar (2015), os autores concluíram que a autoavaliação não foi confiável para avaliar apresentações orais. Em Potter (2020), não foi realizada uma discussão satisfatória acerca dos impactos da autoavaliação por conta de os estudantes terem superestimado demais suas habilidades nos quesitos humanos e subestimado nos quesitos técnicos.

Na pesquisa de Seufert, Guggemos e Sonderegger (2019), os autores questionaram os estudantes sobre suas percepções em relação à autoavaliação após o término do curso. Desrochers et al. (2019) preferiram não discutir os impactos da autoavaliação de seu experimento ainda, e comentaram que é necessário aplicá-lo mais vezes.

3.5.5 QSP05- Acurácia da autoavaliação

A quinta questão secundária de pesquisa preocupou-se em investigar como os autores entendem a acurácia da autoavaliação e como ela foi medida. A Tabela 6 reúne os dados coletados sobre a acurácia:

Tabela 6: Medição da acurácia da autoavaliação.

Medição da acurácia	Quantidade de trabalhos
Não mencionada	21
Correlação de Pearson	12
Comparação com outras estratégias de avaliação	07
Resposta correta às questões técnicas	02
Comparação com exame final	02
Coefficiente de Cohen Kappa	01
<i>Multifaceted Rasch Measurement</i>	01

Na pesquisa de Heukelman (2017), a acurácia da autoavaliação não é discutida, no entanto, os autores mencionam que a autoavaliação revelou um estado mais verdadeiro sobre os conhecimentos dos estudantes do que os testes práticos. Em El-Maaddawy (2017), o autor coloca que a partir da acurácia da autoavaliação, o professor poderia mudar a nota final do estudante. No entanto, não é mencionado como o professor entende uma autoavaliação acurada. Para González-Betancor, Cruz e Verano-Tacoronte (2017), mais estudos são necessários para se ter uma autoavaliação acurada, mas apontam que a falta de intimidade com a autoavaliação, a não participação discente nos critérios de avaliação e o fato dela fazer parte da nota final, fazem com que a acurácia seja prejudicada.

Em Hale (2015), foi combinada uma calibração que ajustaria a nota final quando a divergência entre a avaliação do professor e a autoavaliação resultava em mais de três pontos. Já na pesquisa de Stan e Manea (2015), os autores consideraram a acurácia diretamente relacionada ao quão parecida ou aproximada a autoavaliação foi em relação à avaliação do professor, porém, não mencionaram a fórmula de cálculo utilizada nos totais divulgados. De forma semelhante, em Potter (2020), a autora faz menções à acurácia da autoavaliação comparando-a com os resultados da avaliação em pares.

O coeficiente de correlação de Pearson apareceu em doze (26%) das pesquisas sendo utilizado para explorar a relação entre as estratégias de avaliação utilizadas (avaliação em pares, autoavaliação). Em Guest e Riegler (2017), é mencionada a utilização do coeficiente de correlação que possibilitou a observação de que a relação entre a autoavaliação e avaliação do professor foi fraca. Já em Hosein e Harle (2017), pode-se observar que a acurácia da autoavaliação foi associada ao desempenho em matemática, sendo que os estudantes que se autoavaliaram mais precisamente foram os que mais tinham dificuldades.

Na pesquisa de Ganajová et al. (2020) a acurácia da autoavaliação foi pautada no quão as percepções dos estudantes estavam de acordo com seus conhecimentos reais. Em Pereira et al. (2014), é mencionado o Coeficiente de Cohen Kappa como instrumento para fazer as correlações entre avaliação do professor, autoavaliação e avaliação em pares. Os autores disponibilizam uma tabela onde os valores dos coeficientes de correlação são categorizados, sendo: $< 0,00$ = pobre, $0,00 - 0,20$ = pouco, $0,21 - 0,40$ = razoável, $0,41 - 0,60$ = moderado, $0,61 - 0,80$ = substancial e $0,81 - 1,00$ = quase perfeito.

É importante ressaltar sobre o coeficiente de correlação de Pearson, que ele permite avaliar a correlação entre as estratégias de avaliação utilizadas em conjunto com a autoavaliação somente. Partindo-se do princípio de que quanto mais próximas são a autoavaliação e a avaliação do professor (e dos pares quando usada) mais acurada é a autoavaliação, pode-se entender que tal coeficiente pode auxiliar na determinação da acurácia da autoavaliação de forma geral, e por ter

sido o método mais citado como comparativo entre estratégias nos trabalhos relevantes, tornou-se necessário validá-lo como possibilidade. Mas Pearson fornece o mesmo valor de coeficiente para cenários diferentes e em determinada situação resulta em uma divisão por zero (cenários apresentados na seção 4), o que pode inviabilizar seu uso.

3.5.6 QPP - Como a autoavaliação vem sendo aplicada no ensino superior

Tendo em vista as respostas às questões secundárias de pesquisa, pode-se inferir que a autoavaliação vem sendo aplicada no ensino superior de diferentes formas no que diz respeito ao seu objetivo. Alguns autores enxergam a autoavaliação como um pré-teste, uma avaliação que não vale nota, como reforço dos conteúdos para um exame final. Esse tipo de autoavaliação é reconhecida por Boud e Brew (1995), onde descrevem três tipologias de autoavaliação, sendo esse formato chamado de autoavaliação de “interesse técnico”. Os autores ainda mencionam a tipologia de “interesse comunicativo”, onde os estudantes contribuem com os critérios de avaliação. Essa situação é vista em El Maaddawy (2017), onde o professor inicialmente discute os critérios com os estudantes.

É importante que o estudante tenha participação ou pelo menos conhecimento prévio sobre os critérios de sua autoavaliação e que a avaliação do professor e a avaliação em pares, quando também aplicadas, se baseiem nos mesmos critérios. Para Boud e Brew (1995), se professores e estudantes estão usando os mesmos critérios de avaliação, é fácil de se determinar a relação entre a avaliação do professor e a autoavaliação do estudante. A participação do estudante no desenvolvimento dos critérios, para que sejam específicos e bem entendidos, impacta na acurácia de sua autoavaliação, como colocam Panadero, Brown e Strijbos (2016).

Cada estratégia de avaliação pode desempenhar um papel diferente, e juntas podem ser complementares ao processo de ensino-aprendizagem. Em Khonbi e Sadeghi (2012), por exemplo, a avaliação em pares contribuiu mais do que a autoavaliação, já em Hulsman e Vloodt (2014), as autoavaliações foram mais específicas, e em Sande e Murthy (2014), os estudantes preferiram a avaliação do professor em relação à avaliação em pares.

Escalas foram utilizadas como a forma de resposta em 63% dos trabalhos, tendo como destaque a escala Likert, em 35%. No entanto, no trabalho de Guest e Riegler (2017), foi mencionada a preferência por respostas descritivas, inferindo que dessa forma teriam mais veracidade nas respostas. O *feedback*, seja dos pares, do professor ou da própria autoavaliação, é importante pois pode traduzir melhor a avaliação do que somente a pontuação dos critérios. Tal ponto de vista é ressaltado também por Panadero, Brown e Strijbos (2016), onde inferem que talvez seja mais importante que o estudante ressalte os pontos positivos e negativos do seu trabalho do que atribua uma nota para ele.

Dos dezessete (37%) trabalhos que mencionaram o uso de *softwares* para aplicação da autoavaliação, em nenhum deles a análise de dados pôde ser feita em conjunto, e em seis (13%) outras ferramentas, como o Microsoft Excel e o IBM SPSS, foram utilizadas para a análise. Não foi observado em nenhum trabalho o uso de somente uma ferramenta computacional capaz de aplicar a autoavaliação (e em pares e do professor), fornecer dados sobre sua acurácia (considerando as correlações das estratégias de avaliação empregadas) e também fornecer dados comparativos e gráficos entre as avaliações.

Em vinte e um (46%) trabalhos, a autoavaliação foi utilizada em conjunto com outras estratégias de avaliação como avaliação em pares e do professor. Para esses trabalhos, foram relevantes as comparações entre as valorações de cada tipo de avaliação, tanto para ajuste do processo de ensino-aprendizagem, quanto para observar diferentes pontos de vista e principalmente para determinar a acurácia da autoavaliação. A acurácia da autoavaliação muito foi pautada pela sua semelhança com outras estratégias de avaliação. Apesar de 46% dos trabalhos não dissertarem sobre acurácia, os doze trabalhos que usaram correlação de Pearson e os sete

trabalhos que fizeram comparações com as outras estratégias usadas, pautaram a acurácia nessa comparação. A acurácia é um aspecto que por vezes deixa vulnerável a credibilidade da autoavaliação.

4 Coeficiente de acurácia de autoavaliação

Nos procedimentos realizados nessa pesquisa a acurácia da autoavaliação discente foi apontada como algo difícil de se mensurar. Alguns autores utilizaram o coeficiente de Pearson para determinar o grau de correlação entre duas avaliações, como para determinar se a autoavaliação aproximou-se da avaliação do professor, o que é previsto na fundamentação teórica como uma medida aceita de acurácia. Os valores possíveis para esse coeficiente são de -1 a +1, e a interpretação desse intervalo é mostrada na Figura 4.

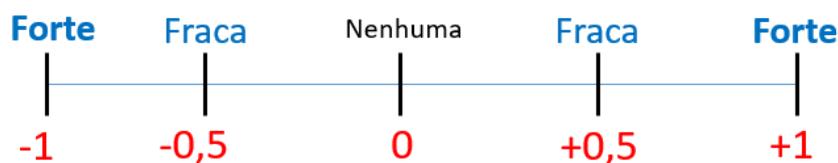


Figura 4: Escala da correlação de Pearson.

No entanto, o Coeficiente de Correlação de Pearson pode não ser determinante em algumas situações. Para exemplificar o motivo da necessidade de um coeficiente de acurácia específico, toma-se como exemplo a rubrica apresentada no Quadro 5 e quatro cenários diferentes: cenário 1, cenário 2, cenário 3 e cenário 4.

Quadro 5: Rubrica de avaliação.

O estudante...	
1= <i>Discordo totalmente</i> 2= <i>Discordo</i> 3= <i>Concordo parcialmente</i> 4= <i>Concordo</i> 5= <i>Concordo totalmente</i>	
Avaliação	
C1- Apresentou coerência na fala	
C2- Demonstrou postura adequada	
C3- Apresentou conteúdo satisfatório	
C4- Envolveu-se na elaboração do trabalho	
C5- Falou por tempo equivalente aos demais	

No cenário 1 apresentado no Quadro 6, as respostas do professor (avaliação do professor) e do estudante (autoavaliação) são mostradas.

Quadro 6: Pontuações do cenário 1.

O estudante "Fulano da Silva"...		
1= <i>Discordo totalmente</i> 2= <i>Discordo</i> 3= <i>Concordo parcialmente</i> 4= <i>Concordo</i> 5= <i>Concordo totalmente</i>		
	Avaliação do professor	Autoavaliação
C1- Apresentou coerência na fala	2	3
C2- Demonstrou postura adequada	2	3
C3- Apresentou conteúdo satisfatório	3	4
C4- Envolveu-se na elaboração do trabalho	2	3
C5- Falou por tempo equivalente aos demais	2	3

No exemplo do Quadro 6, tem-se uma regularidade em relação às respostas do professor e do estudante: o estudante se autoavaliou de maneira geral melhor do que a avaliação do professor, com diferença de um ponto em todos os critérios de avaliação. Ao aplicar o coeficiente de correlação de Pearson nesse cenário, tem-se o valor de coeficiente 1, que indica uma correlação forte e positiva, ou seja, a medida que um valor aumenta, o outro também aumenta e na mesma

proporção. Nessa situação, o valor do coeficiente é o maior possível, o que poderia sugerir que a autoavaliação do estudante está acurada.

Em um segundo cenário, representa-se a mesma constância em relação à autoavaliação ter sempre um valor maior, como pode ser observado no Quadro 7.

Quadro 7: Pontuações do cenário 2.

O estudante “Fulano da Silva”...		
<i>1= Discordo totalmente 2= Discordo 3= Concordo parcialmente 4= Concordo 5= Concordo totalmente</i>		
	Avaliação do professor	Autoavaliação
C1- Apresentou coerência na fala	2	4
C2- Demonstrou postura adequada	2	4
C3- Apresentou conteúdo satisfatório	3	5
C4- Envolveu-se na elaboração do trabalho	2	4
C5- Falou por tempo equivalente aos demais	2	5

No entanto, no cenário 2 há claramente maior discordância entre a autoavaliação e a avaliação do professor, tendo sempre diferença de dois pontos em cada critério de avaliação. Ou seja, no critério “C1- Apresentou coerência na fala” o professor valorou o critério como 2, equivalente ao “discordo”, ao passo que o estudante se autoavalia nesse mesmo critério com a valoração 4 “concordo”.

O valor do coeficiente de correlação de Pearson para esse cenário também é 1, pois o estudante de maneira geral também se autoavaliou melhor que o professor o avaliou, sempre mantendo diferença de dois pontos. Ou seja, a autoavaliação sempre ser maior do que a avaliação do professor e a constante diferença de pontuação assemelha-se a situação apresentada no cenário 1. Porém, em uma rubrica de cinco pontos, dois pontos de diferença são mais significativos do que um ponto de diferença. Sendo assim, o cenário 2 não deveria ser traduzido com o mesmo valor de coeficiente do cenário 1. Já no cenário 3, mostrado no Quadro 8, é apresentada uma situação onde as diferenças de pontuação são mais variantes.

Quadro 8: Pontuações do cenário 3.

O estudante “Fulano da Silva”...		
<i>1= Discordo totalmente 2= Discordo 3= Concordo parcialmente 4= Concordo 5= Concordo totalmente</i>		
	Avaliação do professor	Autoavaliação
C1- Apresentou coerência na fala	3	4
C2- Demonstrou postura adequada	2	3
C3- Apresentou conteúdo satisfatório	3	5
C4- Envolveu-se na elaboração do trabalho	3	3
C5- Falou por tempo equivalente aos demais	4	3

Como pode-se observar, nos critérios C1, C2 e C5, a diferença entre a valoração dada pelo professor e estudante é 1. No entanto, no critério C5, o professor é o indivíduo que atribui a maior pontuação, o que não ocorreu nos cenários anteriores. No critério C4, professor e estudante atribuíram a mesma valoração na avaliação e autoavaliação e por fim, no critério C3 é onde ocorre a maior diferença de pontuação, sendo 2 pontos.

Para esse cenário, o coeficiente de correlação de Pearson é 0, o que indica nenhuma correlação. O resultado do coeficiente é esse, pois, analisando com minúcia, nos cenários anteriores sempre existia a mesma diferença de pontuação e predominância da autoavaliação. Analisando empiricamente o cenário, pode-se inferir que na verdade professor e estudante não estão tendo grandes divergências em suas avaliações, pois inclusive em um dos critérios, ambos atribuíram a mesma pontuação e em outros três, houve diferença de 1 ponto somente, o que significa que essa autoavaliação tem um certo grau de acurácia.

Por fim, no cenário 4, vide Quadro 9, apresenta-se uma situação que o coeficiente de correlação de Pearson não pode determinar: pontuações do estudante e professor exatamente iguais.

Quadro 9: Pontuações do cenário 4.

O estudante “Fulano da Silva”...		
<i>1= Discordo totalmente 2= Discordo 3= Concordo parcialmente 4= Concordo 5= Concordo totalmente</i>		
	Avaliação do professor	Autoavaliação
C1- Apresentou coerência na fala	3	3
C2- Demonstrou postura adequada	3	3
C3- Apresentou conteúdo satisfatório	3	3
C4- Envolveu-se na elaboração do trabalho	3	3
C5- Falou por tempo equivalente aos demais	3	3

Nesse caso a correlação de Pearson não pode ser calculada, pois resulta em uma divisão por zero. No entanto, no cenário 4, avaliação e autoavaliação têm exatamente a mesma visão sobre o estudante, o que deveria ser considerado como 100% de acurácia. Além dessa situação, haja vista que esse coeficiente examina o comportamento da coluna “y” (autoavaliação) de acordo com as alterações da coluna “x” (avaliação do professor), não faz sentido que se esperem resultados diferentes em “y” à medida que “x” permanece igual. No entanto, tratando-se da comparação entre avaliação do professor e autoavaliação de um estudante, esse cenário é possível e seria intratável utilizando esse coeficiente.

Dados os cenários analisados e testados utilizando o coeficiente de correlação de Pearson, descarta-se a utilização desse coeficiente como uma maneira de medir a acurácia de uma autoavaliação em relação à avaliação do professor. Sendo assim, se faz necessária a criação de um coeficiente específico. A próxima seção explicita a fórmula e explicação do coeficiente de acurácia desenvolvido.

4.1 Fórmula e explicação do coeficiente de acurácia

A fórmula do coeficiente de acurácia proposto consiste em:

$$\vartheta = \frac{\sum(ab)}{cd} \quad (2)$$

Considerando: a = total de questões por diferença de pontuação; b = relevância por diferença de pontuação; c = total de questões da rubrica; d = maior relevância (ou *tamanho máximo da escala-1*). O valor final do coeficiente ϑ varia de 0 (acurácia nula) até 1 (totalmente acurada) e pode ser multiplicado por 100 para que se tenha uma representação em porcentagem. A seguir, a fórmula foi aplicada nos quatro cenários mencionados anteriormente.

Haja vista que a correlação de Pearson não leva em consideração as diferenças de pontuação entre as variáveis, a primeira etapa da aplicação da fórmula do coeficiente definido foi a contabilização e atribuição de relevâncias diferentes para cada diferença de pontuação. A contabilização consiste na contagem de quantas vezes a mesma diferença de pontuação aparece na comparação entre autoavaliação e avaliação do professor. Já a atribuição das relevâncias deve ser inversamente proporcional à diferença de pontuação, sendo assim relativa ao tamanho da escala utilizada como resposta na rubrica de avaliação. Essa etapa alimenta a segunda e terceira colunas da “Tabela de quantidades e relevâncias”, apresentada na Tabela 7.

Uma vez realizada a primeira etapa, deve-se construir a tabela que multiplicará a contabilização da diferença de pontuação por sua relevância. A Tabela 7 exemplifica essa construção, tendo como origem de seus dados ainda o cenário 1, apresentado no Quadro 6.

Tabela 7: Tabela de quantidades e relevâncias para o cenário 1.

Diferença de pontuação	Total de critérios com essa diferença (a)	Relevância (b)	Total de critérios com essa diferença vezes relevância
0	0	4	0
1	5	3	15
2	0	2	0
3	0	1	0
4	0	0	0
Soma do total de critérios com essa diferença vezes relevância ($\sum(ab)$)			15

A coluna “Diferença de pontuação” apenas orienta em qual linha devem ser preenchidos os totais das diferenças, e seu valor máximo também depende do tamanho da escala da rubrica e equivale ao resultado do cálculo: *tamanho da escala - 1*. No exemplo do cenário 1, conforme destaca o Quadro 6, todos os cinco critérios ficaram com diferença de 1 ponto, portanto, na linha da diferença “1” a coluna “Total de critérios com essa diferença” foi preenchida com o valor 5. Já a coluna “Relevância” será sempre preenchida com um intervalo inversamente proporcional à “Diferença de pontuação”, ou seja, que parte de *tamanho máximo da escala - 1* e é decrescente. Sendo assim, para uma rubrica com escala de 5 pontos, tem-se como maior relevância o valor 4 para os critérios cujas pontuações de professor e estudante foram iguais (acurácia perfeita), e 0 para a maior diferença de pontuação, que nesse caso é 4 pontos (ex.: professor atribuiu a pontuação 1 a um critério e estudante atribuiu 5 em sua autoavaliação no mesmo critério). Na relevância 0, considera-se a total discordância entre professor e estudante, sendo acurácia nula. Substituindo na fórmula os valores do exemplo do cenário 1, tem-se:

- $\sum(ab)$: $(0 \times 4) + (5 \times 3) + (0 \times 2) + (0 \times 1) + (0 \times 0) = 15$ (vide última linha da Tabela 7);
- c : são 5 questões na rubrica = 5;
- d : 5 (*tamanho máximo da escala*) $- 1 = 4$.

Aplicando na fórmula, tem-se $15 \div (5 \times 4)$, que resulta no valor $\vartheta = 0,75$. Ou seja, para esse cenário, a acurácia é de 75% da autoavaliação discente em relação à avaliação do professor. As Tabelas 8, 9 e 10 exibem os valores correspondentes aos cenários 2, 3 e 4 respectivamente.

Tabela 8: Tabela de quantidades e relevâncias para o cenário 2.

Diferença de pontuação	Total de critérios com essa diferença (a)	Relevância (b)	Total de critérios com essa diferença vezes relevância
0	0	4	0
1	0	3	0
2	5	2	10
3	0	1	0
4	0	0	0
Soma do total de critérios com essa diferença vezes relevância ($\sum(ab)$)			10

Substituindo os valores do cenário 2, apresentado no Quadro 7, na fórmula do coeficiente proposto, tem-se: $10 \div (5 \times 4) = 0,50$. Para esse cenário, a acurácia é de 50%. É importante ressaltar que no coeficiente de correlação de Pearson, os cenários 1 e 2 tinham como resultado o valor de coeficiente 1, implicando a mesma acurácia em ambas as situações. Observa-se então que no coeficiente proposto existe maior coerência no valor resultante, pois em uma escala de 5 pontos, a diferença de 2 pontos na valoração de um critério de avaliação tem uma representação de distância de percepção maior do que a diferença de 1 ponto.

Tabela 9: Tabela de quantidades e relevâncias para o cenário 3.

Diferença de pontuação	Total de critérios com essa diferença (a)	Relevância (b)	Total de critérios com essa diferença vezes relevância
0	1	4	4
1	3	3	9
2	1	2	2
3	0	1	0
4	0	0	0
Soma do total de critérios com essa diferença vezes relevância ($\Sigma(ab)$)			15

Substituindo os valores do cenário 3, apresentado no Quadro 8, na fórmula do coeficiente proposto, tem-se: $15 \div (5 \times 4) = 0,75$. Para esse cenário, a acurácia é de 75%. Na correlação de Pearson, para tal cenário, o resultado do coeficiente foi 0, que segundo a escala apresentada na Figura 4, representa “nenhuma correlação”. No coeficiente proposto, não existe uma distinção em relação à “qual avaliador está avaliando melhor”, seja professor ou estudante. Sendo assim é considerado apenas a quantidade de critérios com determinada diferença de pontuação, o que gera um resultado mais coerente. Por fim, no cenário 4, representado no Quadro 9, onde todas as valorações dos critérios de avaliação do professor e autoavaliação discente eram iguais, o preenchimento da tabela de quantidades e relevâncias fica como mostrado na Tabela 10.

Tabela 10: Tabela de quantidades e relevâncias para o cenário 4.

Diferença de pontuação	Total de critérios com essa diferença (a)	Relevância (b)	Total de critérios com essa diferença vezes relevância
0	5	4	20
1	0	3	0
2	0	2	0
3	0	1	0
4	0	0	0
Soma do total de critérios com essa diferença vezes relevância ($\Sigma(ab)$)			20

O coeficiente de correlação de Pearson não se adequou ao cenário 4 tanto pela questão matemática da divisão por zero quanto pelo sentido de sua aplicação, considerando que com os mesmos valores para “x” (avaliação do professor), não se espera diferentes valores em “y” (autoavaliação discente). No entanto, é visível a total concordância e por consequência acurácia de 100% nessa situação, haja vista que professor e estudante estão tendo exatamente o mesmo olhar em relação ao desempenho do estudante. Tal situação é refletida no valor final ϑ do coeficiente proposto, onde substituindo os valores do cenário 4 na fórmula, tem-se como resultado: $20 \div (5 \times 4) = 1$. Destarte, o coeficiente proposto demonstra 100% de acurácia entre as avaliações.

É importante ressaltar que o coeficiente ainda não foi amplamente testado com estudantes do ensino superior utilizando rubricas quadradas. Os testes foram realizados pelos autores reproduzindo os cenários conhecidos e validando então a eficácia e coerência do coeficiente. No link: <https://bit.ly/3B8JlzS> existe uma planilha do Google que automatiza esse coeficiente. No documento, existem planilhas prontas para diferentes escalas (de 2 a 10) para que os pesquisadores interessados possam criar uma cópia, preencher o documento de acordo com as pontuações da autoavaliação e avaliação do estudante e obter acesso ao grau de acurácia da autoavaliação do estudante.

5 Conclusão

O valor da autoavaliação discente deve ser melhor compreendido pelas instituições de ensino superior que pretendem formar profissionais capazes de refletir e autorregular seu próprio trabalho. Como observado nos trabalhos resultantes deste mapeamento, a autoavaliação traz benefícios educacionais e normalmente não é utilizada de maneira isolada. A avaliação em pares e a avaliação do professor estão sendo aplicadas junto à autoavaliação. No entanto, a acurácia da autoavaliação ainda é um fator que faz com que os professores muitas vezes recuem ao utilizá-la como uma estratégia de avaliação em sala de aula.

O coeficiente de acurácia proposto nessa pesquisa propicia que professor e estudante tenham uma valoração que traduza o quão aproximada a percepção do estudante está de si próprio em relação à percepção do professor. Como benefício, essa valoração minimiza a subjetividade da autoavaliação. No entanto, é preciso considerar que somente para uma ferramenta de rubrica quadrada proposta como autoavaliação e avaliação do professor o coeficiente pode ser utilizado. Para ferramentas de avaliação descritivas ou questionários, ou seja, onde as respostas possam ser de diferentes formas, o coeficiente não poderá ser aplicado.

Como recomendação à comunidade de informática na educação, sugere-se o desenvolvimento de um *software* onde seja possível: criar uma rubrica, aplicá-la como avaliação do professor e autoavaliação do estudante, fornecer gráficos comparando a proximidade das pontuações e exibindo a porcentagem de acurácia da autoavaliação em relação à avaliação do professor utilizando o coeficiente proposto nessa pesquisa. Além disso, sugere-se a atualização do mapeamento incluindo termos inerentes ao ensino médio e profissionalizante para que se tenha um panorama da aplicação da autoavaliação discente em outras modalidades de ensino e se descubram então novas lacunas a serem exploradas. Por fim, aplicar o coeficiente de acurácia proposto em diferentes modalidades de ensino e com rubricas de diferentes tamanhos ajudaria a consolidar sua eficiência.

Algumas limitações e ameaças são cabíveis de se mencionar neste trabalho. Uma delas é que os trabalhos considerados foram apenas em inglês e o intervalo dos anos foi de 2008 a 2020. Tais limitações, mesmo que necessárias para a condução de qualquer pesquisa, podem deixar de retratar a realidade mundial do tópico pesquisado. Outro fator é que no mecanismo de busca Springer, houve uma dificuldade na pesquisa pela variação dos termos, tendo que se optar por pesquisar apenas pelos dois termos principais “*self-assessment*” e “*higher education*”. Essa dificuldade pode ter omitido resultados relevantes.

Referências

- Abolfazli, Z. & Sadeghi, K. (2012). The Effect of Assessment Type (Self vs. Peer vs. Teacher) on Iranian University EFL Students' Course Achievement. *Language Testing in Asia*, 2, 47–74. doi: [10.1186/2229-0443-2-4-47](https://doi.org/10.1186/2229-0443-2-4-47). [GS Search]
- Alishah, A. & Dolmaci, M. (2013). The Interface Between Self-Efficacy Concerning the Self Assessment on Students Studying English as a Foreign Language. *Procedia - social and behavioral sciences*, 70, 873-881. doi: [10.1016/j.sbspro.2013.01.133](https://doi.org/10.1016/j.sbspro.2013.01.133). [GS Search]
- Aryadoust, V. (2015). Self- and Peer Assessments of Oral Presentations by First-Year University Students. *Educational Assessment*, 20, 199-225. doi: [10.1080/10627197.2015.1061989](https://doi.org/10.1080/10627197.2015.1061989). [GS Search]
- Bacich, L. & Moran, J. (2018). *Metodologias Ativas para uma Educação Inovadora*. Porto Alegre: Penso.

- Boud, D. (1995). *Enhancing Learning Through Self-Assessment*. London: Routledge.
- Boud, D. & Falchikov, N. (2005). Redesigning Assessment for Learning Beyond Higher Education. *Higher Education in a Changing World. Research and Development in Higher Education*, 28, 34-41. [GS Search]
- Brew, A. & Boud, D. (1995). Teaching and Research: Establishing the Vital Link With Learning. *Higher Education*, 29, 261-273. doi: [10.1007/BF01384493](https://doi.org/10.1007/BF01384493). [GS Search]
- Brown, G., Andrade, H. & Chen, F. (2015). Accuracy in Student Self-Assessment: Directions and Cautions for Research. *Assessment in Education Principles Policy and Practice*, 22, 444-457. doi: [10.1080/0969594X.2014.996523](https://doi.org/10.1080/0969594X.2014.996523). [GS Search]
- Buchinger, D., Cavalcanti, G. & Hounsell, M. (2014). Mecanismos de Busca Acadêmica: uma Análise Quantitativa. *Revista Brasileira de Computação Aplicada*, 6, 108-120. doi: [10.5335/rbca.2014.3452](https://doi.org/10.5335/rbca.2014.3452). [GS Search]
- Dermeval, D., Coelho, J. A. P. de M. & Bittencourt, I. I. (2019). Mapeamento sistemático e revisão sistemática da literatura em informática na educação (pp. 5-6). Porto Alegre: Sociedade Brasileira de Computação. [GS Search]
- Desrochers, M., Zhang, J., Caron, S. & Steinmiller, J. (2019). An Experimental Comparison of the Effect of Teacher versus Self-Evaluation/Self-Reflection Feedback on College Students' Behavioral Observation Skills. *Journal of Behavioral Education*, 28, 258-274. doi: [10.1007/s10864-018-09313-6](https://doi.org/10.1007/s10864-018-09313-6). [GS Search]
- Dlab, H. M., Katic, M. A., & Candrljic, S. (2015). Ensuring Formative Assessment in Course with Online Tests. *10th International Conference on Computer Science & Education (ICCSE)*, 10, 322–327. doi: [10.1109/ICCSE.2015.7250264](https://doi.org/10.1109/ICCSE.2015.7250264). [GS Search]
- Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, Peer and Co-Assessment in Higher Education: A Review. *Studies in Higher Education*, 24, 331-350. doi: [10.1080/03075079912331379935](https://doi.org/10.1080/03075079912331379935). [GS Search]
- Echeazarra, L., Perera, J., Sanz-santamaría, S. & Gutiérrez, J. (2014). Student-Generated Online Videos to Develop Cross-Curricular and Curricular Competencies in Nursing Studies. *Computers in Human Behavior*, 31, 580-590. doi: [10.1016/j.chb.2013.06.011](https://doi.org/10.1016/j.chb.2013.06.011). [GS Search]
- El Maaddawy, T. (2017). Enhancing Learning of Engineering Students Through Self-Assessment. *IEEE Global Engineering Education Conference (EDUCON)*, 86-91. doi: [10.1109/EDUCON.2017.7942828](https://doi.org/10.1109/EDUCON.2017.7942828). [GS Search]
- Ezzahra, K., Islam, O., & Radid, M. (2015). Self-assessment of the Progress of Thesis for the PHD Students in the Moroccan University. *Procedia - Social and Behavioral Sciences*, 197, 1789-1795. doi: [10.1016/j.sbspro.2015.07.237](https://doi.org/10.1016/j.sbspro.2015.07.237). [GS Search]
- França, R. & Tedesco, P. (2015). Caracterizando a Pesquisa Sobre Autoavaliação na Aprendizagem de Programação para Iniciantes. *Simpósio Brasileiro de Informática na Educação*, 549-558. doi: [10.5753/cbie.sbie.2015.549](https://doi.org/10.5753/cbie.sbie.2015.549).
- Ganajová, M., Sotáková, I., Jurková, V., Brestenska, B., Szarka, K. & Kozurkova, M. (2020). Investigating Students' Self-Assessment of Conceptual Understanding Using Self Assessment Cards in Chemistry. *ICMET '20: Proceedings of the 2020 2nd International Conference on Modern Educational Technology*, 20, 30-33. doi: [10.1145/3401861.3401864](https://doi.org/10.1145/3401861.3401864). [GS Search]
- Garson, G. D. (2009). Statnotes: Topics in Multivariate Analysis. Recuperado de: <https://faculty.chass.ncsu.edu/garson/pa765/statnote.htm>. Acesso em: fevereiro de 2022.

- González-Betancor, S. M., Bolívar-Cruz, A. & Verano-Tacoronte, D. (2017). Self-assessment Accuracy in Higher Education: The Influence of Gender and Performance of University Students. *Active Learning in Higher Education*, 20, 101-114. doi: [10.1177/1469787417735604](https://doi.org/10.1177/1469787417735604). [GS Search]
- Guest, J. & Riegler, R. (2016). Learning by Doing: Do Economics Students Self-Evaluation Skills Improve? *International Review of Economics Education*, 24, 50-64. doi: [10.1016/j.iree.2016.10.002](https://doi.org/10.1016/j.iree.2016.10.002). [GS Search]
- Hale C. (2015). Self-Assessment as Academic Community Building: a Study from a Japanese Liberal Arts University. *Language Testing in Asia*, 5, 1-12. doi: [10.1186/s40468-014-0010-0](https://doi.org/10.1186/s40468-014-0010-0). [GS Search]
- Han, C. (2018). A Longitudinal Quantitative Investigation Into the Concurrent Validity of Self and Peer Assessment Applied to English-Chinese Bi-Directional Interpretation in an Undergraduate Interpreting Course. *Studies in Educational Evaluation*, 58, 187-196. doi: [10.1016/j.stueduc.2018.01.001](https://doi.org/10.1016/j.stueduc.2018.01.001). [GS Search]
- Heukelman, D. (2017). Measuring E-Readiness: A Case Study: Self-Assessment vs Performance Testing. *1st International Conference on Next Generation Computing Applications (NextComp)*, 215–219. doi: [10.1109/NEXTCOMP.2017.8016201](https://doi.org/10.1109/NEXTCOMP.2017.8016201). [GS Search]
- Hosein, A. & Harle, J. (2018). The Relationship Between Students' Prior Mathematical Attainment, Knowledge and Confidence on Their Self-Assessment Accuracy. *Studies in Educational Evaluation*, 56, 32–41. doi: [10.1016/j.stueduc.2017.10.008](https://doi.org/10.1016/j.stueduc.2017.10.008). [GS Search]
- Hulsman, R. & Vloodt, J. (2014). Self-Evaluation and Peer-Feedback of Medical Students' Communication Skills Using a Web-Based Video Annotation System. Exploring Content and Specificity. *Patient Education and Counseling*, 98. doi: [10.1016/j.pec.2014.11.007](https://doi.org/10.1016/j.pec.2014.11.007). [GS Search]
- Hwang, W., Hsu, J., Shadiev, R., Chang, C. & Huang, Y. M. (2015). Employing Self-Assessment, Journaling, and Peer Sharing to Enhance Learning From an Online Course. *Journal of Computing in Higher Education*, 27, 114-133. doi: [10.1007/s12528-015-9096-3](https://doi.org/10.1007/s12528-015-9096-3). [GS Search]
- Ibabe, I. & Jauregizar, J. (2010). Online Self-Assessment With Feedback and Metacognitive Knowledge. *Higher Education*, 59, 243–258. doi: [10.1007/s10734-009-9245-6](https://doi.org/10.1007/s10734-009-9245-6). [GS Search]
- Kitchenham, B. & Charters S. (2007). Guidelines for performing systematic literature reviews in software engineering. Keele: Software Engineering Group - School of Computer Science and Mathematics. [GS Search]
- Lanthony, A., Azzouzi, E., Francois, A. & Peyret, N. (2018). Self-Assessment In PBL: A Tool to Develop Self-Confidence and Autonomy of Students: The Example of Self-Assessment Experiment At Ismep-Supmeca In France. *2018 IEEE Global Engineering Education Conference (EDUCON)*, 263-267. doi: [10.1109/Educon.2018.8363238](https://doi.org/10.1109/Educon.2018.8363238). [GS Search]
- Larson, R. & Farber, B. (2015). *Estatística aplicada*. São Paulo: Pearson Education do Brasil.
- Lewis, M., Titterton, N. & Clancy, M. (2011). Developing Students' Self-Assessment Skills Using Lab-Centric Instruction. *Journal Of Computing Sciences In Colleges*, 26, 173-180. doi: [10.5555/1953573.1953602](https://doi.org/10.5555/1953573.1953602). [GS Search]
- Liang, Y., Wang, Q., Lu, Y., Qian, R. & Yin, Y. (2011). Using a Web-Based System to Explore Peer, Self, and Tutor Assessment in Problem-Based Learning Tutorials. *ITME 2011 - Proceedings: 2011 IEEE International Symposium On IT In Medicine And Education*, 1, 253-256. doi: [10.1109/ITiME.2011.6130827](https://doi.org/10.1109/ITiME.2011.6130827). [GS Search]

- Mahlberg, J. (2015). Formative Self-Assessment College Classes Improves Self-Regulation And Retention In First/Second Year Community College Students. *Community College Journal of Research and Practice*, 39, 772-783. doi: [10.1080/10668926.2014.922134](https://doi.org/10.1080/10668926.2014.922134). [GS Search]
- Matsuno, S. (2009). Self, Peer-, and Teacher-Assessments in Japanese University EFL Writing Classrooms. *Language Testing - Lang Test*, 26, 75-100. doi: [10.1177/0265532208097337](https://doi.org/10.1177/0265532208097337). [GS Search]
- Mbonigaba, J. & Oumar, S. (2015). Exploring the Reliability of Self-Assessment and Peer Assessment in Oral Presentations in Economics: A Sample of Postgraduate Students at a South African University. *International Journal of Educational Sciences*, 9, 81-92. doi: [10.1080/09751122.2015.11890297](https://doi.org/10.1080/09751122.2015.11890297). [GS Search]
- Mendoza, B., Reyes-Alamo, J., Wu, H., Carranza, A. & Zavala, L. (2016). iPractice: A Self-Assessment Tool for Students Learning Computer Programming in an Urban Campus. *Journal of Computing Sciences in Colleges*, 31, 93-100. doi: [10.5555/2835377.2835392](https://doi.org/10.5555/2835377.2835392). [GS Search]
- Meng, C., Cui, G. & Wang, S. (2015). The Effects of Self-Evaluation Traits on Efl Students' Learning Satisfaction. *2015 International Conference of Educational Innovation Through Technology (EITT)*, 184-189. doi: [10.1109/Eitt.2015.46](https://doi.org/10.1109/Eitt.2015.46). [GS Search]
- Montalvo, T. & Torres, M. C. (2004). El Aprendizaje Autoregulado: Presente y Future de la Investigación. *Electronic Journal Of Research In Educational Psychology*, 2, 1-33. doi: [10.25115/ejrep.3.120](https://doi.org/10.25115/ejrep.3.120). [GS Search]
- Monteiro, E. J. L. B., Leitão, G. de S. & Barreto, R. da S. (2021). Um Método Baseado na Teoria da Resposta ao Item para Avaliação e Feedback Automático no Contexto de Educação Digital. *Revista Brasileira de Informática na Educação*, 29, 746-774. doi: [10.5753/rbie.2021.29.0.746](https://doi.org/10.5753/rbie.2021.29.0.746). [GS Search]
- Oliveira, K. L. & Santos, A. (2005). Avaliação de Aprendizagem na Universidade. *Psicologia Escolar e Educacional*, 9, 37-46. doi: [10.1590/S1413-85572005000100004](https://doi.org/10.1590/S1413-85572005000100004). [GS Search]
- Othman, W. & Abdullah, A. (2018). Assessing Self-Efficacy and College Readiness Level Among New Undergraduate Students in Computer Science Using Metacognitive Awareness Inventory (MAI). *IOP Conference Series: Materials Science And Engineering*. doi: [10.1088/1757-899x/342/1/012067](https://doi.org/10.1088/1757-899x/342/1/012067). [GS Search]
- Panadero, E., Brown, G. & Strijbos, J. (2016). The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions. *Educational Psychology Review*, 28, 803-830. doi: [10.1007/s10648-015-9350-2](https://doi.org/10.1007/s10648-015-9350-2). [GS Search]
- Perera, J., Mohamadou, G. & Kaur, S. (2009). The Use of Objective Structured Self-Assessment and Peer-Feedback (OSSP) For Learning Communication Skills: Evaluation Using a Controlled Trial. *Advances in Health Sciences Education: Theory and Practice*, 15, 185–193. doi: [10.1007/S10459-009-9191-1](https://doi.org/10.1007/S10459-009-9191-1). [GS Search]
- Petersen, K., Vakkalanka, S. & Kuzniarz, L. (2015). Guidelines for Conducting Systematic Mapping Studies in Software Engineering: An Update. *Information and Software Technology*, 64, 1-18. doi: [10.1016/j.infsof.2015.03.007](https://doi.org/10.1016/j.infsof.2015.03.007). [GS Search]
- Pop, C. & Khampirat, B. (2019). Self-Assessment Instrument to Measure the Competencies of Namibian Graduates: Testing of Validity and Reliability. *Studies in Educational Evaluation*, 60, 130-139. doi: [10.1016/J.Stueduc.2018.12.004](https://doi.org/10.1016/J.Stueduc.2018.12.004). [GS Search]

- Potter, L. E. (2020). Preparing for Projects: IT Student Self-Evaluation of Technical and Professional Skills. *Proceedings of the 2020 on Computers and People Research Conference*, 63-69. doi: [10.1145/3378539.3393868](https://doi.org/10.1145/3378539.3393868). [GS Search]
- Ro, H. K., Lattuca, L. & Alcott, B. (2017). Who Goes to Graduate School? Engineers' Math Proficiency, College Experience, and Self-Assessment of Skills: Choice of Engineering Graduate School. *Journal of Engineering Education*, 106, 98-122. doi: [10.1002/Jee.20154](https://doi.org/10.1002/Jee.20154). [GS Search]
- Sánchez, V. G., Soldado, R. M. & López, M. C. P. (2014). Self-assessment via a blended-Learning Strategy to Improve Performance in an Accounting Subject. *International Journal of Educational Technology in Higher Education*, 11, 43-54. doi: [10.7238/rusc.v11i2.2055](https://doi.org/10.7238/rusc.v11i2.2055). [GS Search]
- Sande, J.C. & Murthy, A. (2014). Including Peer and Self-Assessment in a Continuous Assessment Scheme in Electrical and Electronics Engineering Courses. *Proceedings - Frontiers in Education Conference*, 1-5. doi: [10.1109/FIE.2014.7044284](https://doi.org/10.1109/FIE.2014.7044284). [GS Search]
- Serra-Toro, C., Traver, V. & Amengual, J. (2015). Promoting Student Commitment and Responsibility Through Self- And Peer-Assessment. *Proceedings - Frontiers In Education Conference*, 1-4. doi: [10.1109/FIE.2014.7044285](https://doi.org/10.1109/FIE.2014.7044285). [GS Search]
- Seufert, S., Guggemos, J. & Sonderegger, S. (2019). Learning Analytics in Higher Education using Peer-feedback and Self-assessment: Use Case of an Academic Writing Course. *Proceedings of the 11th International Conference on Computer Supported Education*, 2, 315-322. doi: [10.5220/0007714603150322](https://doi.org/10.5220/0007714603150322). [GS Search]
- Silva Junior, S. D. & Costa, F. J. (2014). Mensuração e Escalas de Verificação uma Análise Comparativa das Escalas Likert e Phase Completion. XVII SEMEAD *Seminários em Administração* (pp. 4-5). [GS Search]
- Spencer, L., Marschark, M., Machmer, E., Durkin, A., Borgna, G. & Convertino, C. (2018). Communication Skills of Deaf And Hard-Of-Hearing College Students: Objective Measures and Self-Assessment. *Journal of Communication Disorders*, 75, 13-24. doi: [10.1016/J.Jcomdis.2018.06.002](https://doi.org/10.1016/J.Jcomdis.2018.06.002). [GS Search]
- Ssegawa, J. K. & Kasule, D. (2017). A Self-Assessment of the Propensity to Obtain Future Employment: A Case of Final-Year Engineering Students at The University of Botswana. *European Journal of Engineering Education*, 42, 513-532. doi: [10.1080/03043797.2016.1193124](https://doi.org/10.1080/03043797.2016.1193124). [GS Search]
- Stan, C. & Manea, A. (2015). The Divergent Relationship Between Assessment and SelfAssessment in Higher Education. Experimental Results. *Procedia - Social and Behavioral Sciences*, 209, 497-502. doi: [10.1016/j.sbspro.2015.11.278](https://doi.org/10.1016/j.sbspro.2015.11.278). [GS Search]
- Tan, K. H. K. (2008). Qualitatively Different Ways of Experiencing Student Self-Assessment. *Higher Education Research & Development*, 27, 15-29. doi: [10.1080/07294360701658708](https://doi.org/10.1080/07294360701658708). [GS Search]
- Torres-Guijarro, S. & Bengoechea, M. (2016). Gender Differential in Self-Assessment: A Fact Neglected in Higher Education Peer and Self-Assessment Techniques. *Higher Education Research and Development*, 36, 1072-1084. doi: [10.1080/07294360.2016.1264372](https://doi.org/10.1080/07294360.2016.1264372). [GS Search]
- Vieira, F. & Moreira, M. A. (2003). Para além dos testes: A avaliação processual na aula de Inglês. Braga: Universidade Do Minho.

- Vieira, I. M. A. (2013). A autoavaliação como instrumento de regulação da aprendizagem (pp. 28). Lisboa: Universidade Aberta. Departamento de Educação e Ensino a Distância. [GS Search]
- Walser, T. (2009). An Action Research Study of Student Self-Assessment in Higher Education. *Innovative Higher Education*, 34, 299-306. doi: [10.1007/S10755-009-9116-1](https://doi.org/10.1007/S10755-009-9116-1). [GS Search]
- Wegner, S. & Beumann, S. (2017). An Outlook on Self-Assessment of Homework Assignments in Higher Mathematics Education. *International Journal of STEM Education*, 5, 55-62. doi: [10.1186/S40594-018-0146-Z](https://doi.org/10.1186/S40594-018-0146-Z). [GS Search]
- Weisi, H. & Karimi, M. (2013). The Effect of Self-Assessment Among Iranian EFL Learners. *Procedia - Social and Behavioral Sciences*, 70, 731-737. doi: [10.1016/J.Sbspro.2013.01.117](https://doi.org/10.1016/J.Sbspro.2013.01.117). [GS Search]

Apêndices

Apêndice A: Trabalhos relevantes

Ano	Autor(es)	Título	País
2009	Walser	<i>An Action Research Study of Student Self-Assessment in Higher Education</i>	Estados Unidos
2009	Perera, Mohamadou e Kaur	<i>The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: evaluation using a controlled trial</i>	Malásia
2009	Matsuno	<i>Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms</i>	Japão
2010	Ibabe e Jauregizar	<i>Online self-assessment with feedback and metacognitive knowledge</i>	Espanha
2011	Lewis, Titterton e Clancy	<i>Developing students self-assessment skills using lab-centric instruction</i>	Estados Unidos
2011	Yan et al.	<i>Using a web-based system to explore peer, self, and tutor assessment in problem-based learning tutori</i>	China
2012	Weisi e Karimi	<i>The effect of self - assessment among Iranian EFL learners</i>	Irã
2012	Alishah e Dolmaci	<i>The interface between self-efficacy concerning the self-assessment on students studying English as a foreign language</i>	Turquia
2012	Khonbi e Sadeghi	<i>The effect of assessment type (self vs. peer) on Iranian university EFL students course achievement</i>	Irã
2014	Sánchez, Soldado e López	<i>Self-assessment via a blended-learning strategy to improve performance in an accounting subject</i>	Espanha
<i>(continua na próxima página)</i>			
2014	Pereira et al.	<i>Student-generated online videos to develop cross-curricular and curricular competencies in Nursing Studies</i>	Espanha
2014	Toro, Traver e Amengual	<i>Promoting student commitment and responsibility through self- and peer-assessment</i>	Espanha
2014	Sande e Murthy	<i>Including Peer and Self-Assessment in a Continuous Assessment Scheme in Electrical and Electronics Engineering Courses</i>	Espanha
2015	Mendoza et al.	<i>Ipractice: a self-assessment tool for students learning computer programming in an urban campus</i>	Estados Unidos

2015	Hale	<i>Self-assessment as academic community building: a study from a Japanese liberal arts university</i>	Japão
2015	Diogo, Barbosa e Ferreira	<i>A pilot Tuning Project-based national study on recently graduated medical students' self-assessment of competences - the TEST study</i>	Portugal
2015	Hwang et al.	<i>Employing self-assessment, journaling, and peer sharing to enhance learning from an online course</i>	Estados Unidos
2015	Hulsman e Vloodt	<i>Self-evaluation and peer-feedback of medical students' communication skills using a web-based video annotation system. Exploring content and specificity</i>	Holanda
2015	Dlab, Katic, Candric	<i>Ensuring Formative Assessment in E-Course with Online Tests</i>	Croácia
2015	Meng, Cui e Wang	<i>The Effects of Self-evaluation Traits on EFL Students' Learning Satisfaction</i>	China
2015	Ezzahra, Islam e Radid	<i>Self-assessment of the progress of thesis for the phd students in the Moroccan university</i>	Marrocos
2015	Stan e Manea	<i>The divergent relationship between assessment and self-assessment in higher education. Experimental results</i>	Romênia
2015	Mahlberg	<i>Formative Self-Assessment College Classes Improves Self-Regulation and Retention in First/Second Year Community College Students</i>	Estados Unidos
2015	Aryadoust	<i>Self- and Peer Assessments of Oral Presentations by First-Year University Students</i>	Singapura
2015	Mbonigaba e Oumar	<i>Exploring the Reliability of Self-Assessment and Peer-Assessment in Oral Presentations in Economics: A Sample of Postgraduate Students at a South African University</i>	Camarões
2016	Guijarro e Bengoechea	<i>Gender differential in self-assessment: a fact neglected in higher education peer and selfassessment techniques</i>	Espanha
2017	Ro, Lattuca e Alcott	<i>Who Goes to Graduate School? Engineers' Math Proficiency, College Experience, and Self-Assessment of Skills</i>	Estados Unidos
2017	Ssegawa e Kasule	<i>A self-assessment of the propensity to obtain future employment: a case of final-year engineering students at the University of Botswana</i>	Botsuana
2017	Guest e Riegler	<i>Learning by doing: Do economics students self-evaluation skills improve?</i>	Inglaterra
2017	Heukelman	<i>Measuring e-Readiness: A case study Self-assessment vs Performance Testing</i>	África do Sul
<i>(continua na próxima página)</i>			
2017	El-Maaddawy	<i>Enhancing Learning of Engineering Students Through Self-Assessment</i>	Emirados Árabes
2017	González-Betancor, Cruz e Verano-Tacoronte	<i>Self-assessment accuracy in higher education: The influence of gender and performance of university students</i>	Espanha
2018	Othman e Abdullah	<i>Assessing self-efficacy and college readiness level among new undergraduate students in computer science using metacognitive awareness inventory (MAI)</i>	Malásia

2018	Han	<i>A longitudinal quantitative investigation into the concurrent validity of self and peer assessment applied to English-Chinese bi-directional interpretation in an undergraduate interpreting course</i>	China
2018	Hosein e Harle	<i>The relationship between students' prior mathematical attainment, knowledge and confidence on their self-assessment accuracy</i>	Inglaterra
2018	Lanthyony et al.	<i>Self-assessment in PBL: a tool to develop selfconfidence and autonomy of students: The example of self-assessment experiment at ISMEP-Supmeca in France</i>	Espanha
2018	Spencer et al.	<i>Communication skills of deaf and hard-of-hearing college students: Objective measures and self-assessment</i>	Estados Unidos
2018	Beumann e Wegner	<i>An outlook on self-assessment of homework assignments in higher mathematics education</i>	Inglaterra
2019	Desrochers et al.	<i>An Experimental Comparison of the Effect of Teacher Versus Self-Evaluation/Self-Reflection Feedback on College Students' Behavioral Observation Skills</i>	Estados Unidos
2019	Pop e Khampirat	<i>Self-assessment instrument to measure the competencies of Namibian graduates: Testing of validity and reliability</i>	Tailândia
2019	Duran et al.	<i>Exploring the Value of Student Self-Evaluation in Introductory Programming</i>	Canadá
2019	Seufert, Guggemos e Sonderegger	<i>Learning Analytics in Higher Education using Peer-feedback and Self-assessment: Use Case of an Academic Writing Course</i>	Suíça
2019	Hung et al.	<i>Bridging assessment and achievement: Repeated practice of self-assessment in college english classes in Taiwan</i>	China
2020	Potter	<i>Preparing for Projects: IT Student Self-Evaluation of Technical and Professional Skills</i>	Austrália
2020	Ganajová et al.	<i>Investigating Students' Self-Assessment of Conceptual Understanding Using Self-Assessment Cards in Chemistry</i>	Eslováquia
2020	Lladó et al.	<i>An analysis of teamwork based on self and peer evaluation in higher education</i>	Espanha