

AFP-OPA: Arcabouço para a Escolha de Pares na Online Peer Assessment

Leonam Cordeiro de Oliveira¹, Selma dos Santos Rosa², Andrey Ricardo Pimentel¹

¹Departamento de Informática – Universidade Federal do Paraná (UFPR)
Curitiba – PR – Brasil

²Centro de Estudos do Mar – Universidade Federal do Paraná (UFPR)
Pontal do Paraná – PR – Brasil

leonamcdeoliveira@gmail.com, andrey@inf.ufpr.br, selmasantos@ufpr.br

Abstract. *Peer Assessment is an alternative assessment method in which students evaluate the work of their colleagues. There are reports from students about feeling uncomfortable about evaluating and being evaluated by peers. A framework has been proposed for the formation of peers based on student characteristics and also a module for detecting possible evaluations poorly performed by students, in support to the teacher. An experiment with 63 students was carried out and initial results indicate that the pairs formed from characteristics can be a way to mitigate these discomforts, as well as indicated a learning gain at the time they perform the evaluations, instead of after receiving feedback activities.*

Resumo. *Peer Assessment é um método de avaliação alternativo no qual os estudantes avaliam os trabalhos de seus colegas. Há relatos de estudantes sobre sentirem desconforto de avaliar e ser avaliado pelos pares. Foi proposto um arcabouço para a formação dos pares a partir de características dos estudantes e também um módulo de detecção de possíveis avaliações mal realizadas pelos estudantes, como apoio ao docente. Um experimento com 63 estudantes foi realizado e resultados iniciais indicam que os pares formados a partir de características pode ser uma maneira de mitigar estes desconfortos, assim como indicou um ganho de aprendizagem no momento que realizam as avaliações, ao invés de após o recebimento do feedback de suas atividades.*

1. Introdução

Tem-se buscado, com o passar dos anos, uma aprendizagem mais centrada no estudante e que também utilize estratégias para a adoção da colaboração entre os estudantes [Pereira et al., 2010]. É importante considerar, também, que é usual que estudantes tentem resolver seus problemas inicialmente com seus colegas, antes de buscarem o auxílio do docente [Boud et al., 2014].

Dentro da área de informática na educação, a aprendizagem colaborativa com suporte computacional tem sido utilizada para propor o desenvolvimento de *software* e aplicações que propiciem a aprendizagem em grupo, oferecendo atividades criativas que explorem o intelecto e a interação social [Stahl et al., 2006]. Além da socialização entre

estudantes no processo de ensino, há ainda a possibilidade de que a etapa da avaliação da aprendizagem também seja colaborativa.

Conforme [Fernandes, 2005], busca-se uma avaliação alternativa mais ambientada em contextos vividos por docentes e estudantes, mais participativa e com maior integração dos processos de ensino e de aprendizagem. Um exemplo de avaliação alternativa que busca a colaboração entre os estudantes é a *Online Peer Assessment* (OPA), ou avaliação *online* por pares, em português.

A OPA surge a partir da *Peer Assessment*, que é a avaliação em que estudantes irão especificar “o nível, a qualidade e o valor de um produto ou o desempenho de outros alunos com igual status” [Topping, 2009, p. 20]. Em outras palavras, a *Peer Assessment* propõe que estudantes avaliem as atividades e trabalhos de seus colegas de turma, buscando a partilha de conhecimentos, ideias e experiências [Boud et al., 2014].

Já a OPA utiliza tecnologia computacional para a sua realização, ajudando na construção de estratégias de avaliação e também facilitando a troca de *feedback* entre pares [Rosa et al., 2017]. É importante que esse *feedback* seja bem utilizado, para que a motivação do ensino e da aprendizagem, com base no próprio processo de avaliação, seja potencializada [Pereira et al., 2010].

Alguns desafios ainda são encontrados quando a OPA é proposta em sala de aula. Entre os relatos dos discentes, é pontuado que há desconforto ao avaliar seus colegas [Chen, 2010] e também ao serem avaliados pelos seus colegas [Wen e Tsai, 2008]. A partir de uma revisão sistemática da literatura, foi possível perceber que, entre os processos da OPA que estão sendo estudados, não há estudos que referenciam a formação dos pares neste processo, sendo que a maioria dos pares são escolhidos de forma aleatória ou então manualmente pelo próprio docente [Oliveira, 2019].

A partir dos problemas citados acima, esta pesquisa teve como objetivo propor uma solução computacional para reduzir o desconforto do estudante e melhorar sua participação nos processos da *Online Peer Assessment*. A OPA tem sido tema de estudos atualmente e, dentre os trabalhos relacionados, destacam-se as pesquisas de [Rosa et al., 2017] e [Alves et al., 2018], publicadas em eventos e periódicos brasileiros que discutem a avaliação nas perspectivas da educação e informática na educação.

A partir de uma revisão sistemática da literatura, [Rosa et al., 2017] pontuam a importância da tecnologia computacional na realização de avaliações. As autoras citam como desafio para estudos futuros o de compreender como as tecnologias de informação e comunicação podem apoiar às práticas de avaliação, tanto na prática da OPA com tecnologias já existentes, quanto no desenvolvimento de recursos específicos para a educação.

Já trabalhos como o de [Alves et al., 2018] realizam o desenvolvimento e a aplicação efetiva de ferramentas que auxiliam na avaliação realizada por pares. Na ferramenta proposta, é realizada a análise de correlação entre as avaliações realizadas pelos estudantes e pelo docente, o que permite dar apoio aos docentes na realização de avaliações da aprendizagem e na compreensão do comportamento discente.

Seguindo o caminho que estas pesquisas apontam e do objetivo de pesquisa proposto, na seção 2 a tecnologia computacional desenvolvida para esta pesquisa é

apresentada e discutida. Na sequência, na seção 3, a metodologia e as perguntas de pesquisa são apresentadas. A seção 4 apresenta os resultados e discussões desta pesquisa e na seção 5 são apresentados os trabalhos futuros e as considerações finais.

2. AFP-OPA

O AFP-OPA, acrônimo de Arcabouço de Formação de Pares para *Online Peer Assessment*, consiste em dois módulos desenvolvidos (Figura 1): o módulo do estudante e o módulo do professor. Todos os atributos e características definidos e discutidos a seguir, relacionados com o AFP-OPA, partem da realização de uma revisão sistemática da literatura sobre aprendizagem colaborativa com suporte computacional e pode ser conferida em [Oliveira et al., 2019a].

É importante destacar que o AFP-OPA, conforme mostra a Figura 1, está inserido em um *software* desenvolvido para a realização da *Online Peer Assessment*, chamado de “OPA!”¹. Vale pontuar que o *software* já dispunha previamente das possibilidades de formação dos pares de forma aleatória e manual pelo docente.

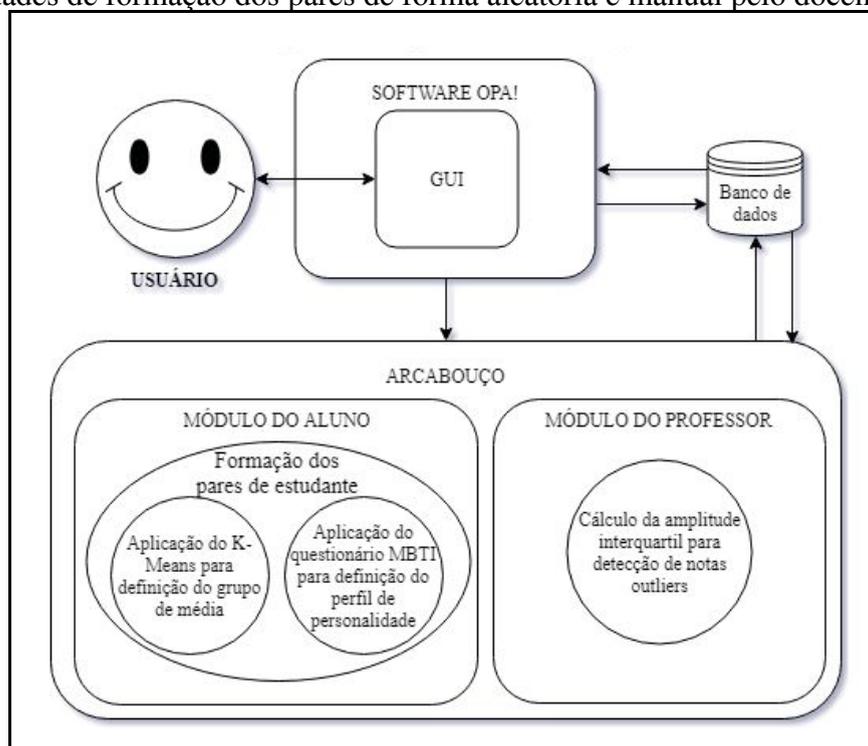


Figura 1 - AFP-OPA. Adaptado de [Oliveira, 2019].

O módulo do estudante foi previamente apresentado em [Oliveira et al., 2019b] e consiste no módulo de formação dos pares dos estudantes, que leva em conta dois atributos dos estudantes. O primeiro atributo são as notas prévias do estudante na disciplina (como atividades, provas e trabalhos práticos) que foram agrupadas em quatro grupos de similaridade, utilizando o algoritmo *k-means*.

O segundo atributo é o perfil de personalidade MBTI do estudante, definido nos estudos de [Briggs-Myers et al., 2003], que é um indicador de tipos psicológicos que

¹ Link para acesso ao OPA!: <http://softwareopa.flisoljs.info>.

busca auxiliar as pessoas a identificarem suas preferências pessoais mais significativas a partir de um questionário [Couto et al., 2016]. O MBTI prevê 16 possíveis personalidades que são apresentadas por siglas, conforme visualizado na parte inferior esquerda da Figura 2, sendo que os tipos de personalidade podem ser divididos em quatro grupo de similaridade, seguindo o protocolo do Keirsey Temperament Sorter II [Keirsey e Bates, 1998].

Definiu-se, para a primeira versão do AFP-OPA, que cada par seria avaliado por quatro pares avaliadores, seguindo às recomendações de [Bouzidi e Jaillet, 2009], que indica que pelo menos quatro avaliações para cada estudante são necessárias para que a avaliação seja relevante. A partir das informações coletadas, busca-se formar o grupo de pares avaliadores de cada estudante o mais heterogêneo possível.

Para o desenvolvimento do modelo do estudante foi necessário agrupar os dados dos estudantes em quatro grupos distintos, conforme pode ser visualizado na parte superior da Figura 2, onde em cada grupo estavam inseridos quatro *arrays*. Cada um dos 16 *arrays* que foram criados estavam relacionados a um tipo de personalidade MBTI e a um grupo de média de notas e cada estudante era associado a um destes *arrays*.

A partir dessas informações, cada estudante, no papel de par avaliado, era associado a quatro pares avaliadores em uma busca nestes *arrays*. Inicialmente buscava-se um par avaliador de cada um dos quatro grupos e quando não era possível utilizava-se a distância euclidiana, que é o cálculo da distância entre dois pontos, para encontrar o par mais distante possível do par avaliado.

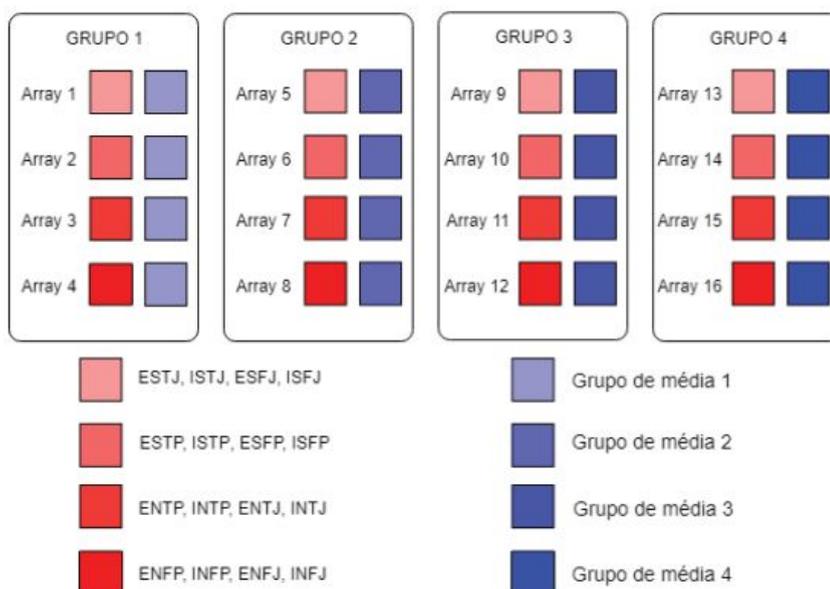


Figura 2 - Modelo de agrupamento dos estudantes. Adaptado de [Oliveira, 2019].

No módulo do professor, buscou-se detectar possíveis notas recebidas pelos estudantes que pudessem estar fora do contexto esperado, a partir de comparação com as outras notas recebidas pelo mesmo par avaliado. Para isso utilizou-se o cálculo da amplitude interquartil que busca identificar, a partir de um conjunto de dados, possíveis dados que fujam do padrão dos demais, chamados *outliers* [Chandola et al., 2007].

3. Metodologia

A partir do objetivo do estudo, de desenvolver o AFP-OPA para diminuir o desconforto dos estudantes e melhorar sua participação nos processos da OPA, foram definidas duas perguntas de pesquisa (PP) para este estudo, conforme apresentado previamente em [Oliveira et al., 2019b]:

- PP1: Formar pares heterogêneos a partir de atributos e características específicas do estudante pode melhorar a sua confiança, diminuindo seu desconforto no processo de OPA?
- PP2: Considerar atributos e características do estudantes na formação de pares avaliadores para a OPA contribui para o desempenho acadêmico, sob a perspectiva discente, quanto à atividade avaliada?

Para isso, portanto, foi proposto um experimento em sala de aula, com turmas da disciplina Introdução a Informática do curso de Automação Mecânica de uma instituição pública de ensino superior. Duas turmas de um mesmo docente foram disponibilizadas para a realização do experimento, contendo um total de 63 estudantes participantes na pesquisa e 1 docente.

Os estudantes receberam uma explicação referente ao experimento que iriam realizar e assinaram um termo de consentimento livre e esclarecido, sendo que não foram informados sobre a estratégia utilizada para a formação dos pares: de forma aleatória, formando o grupo de controle, ou de forma automática pelo AFP-OPA, formando o grupo experimental.

Foi solicitado a todos os estudantes que respondessem um questionário de personalidade MBTI para ser inserido no sistema e em seguida que realizassem uma atividade de descrição narrativa de uma situação proposta pelo professor. Toda a atividade foi realizada utilizando o *software* “OPA!”, tanto o desenvolvimento da descrição narrativa quanto a avaliação das atividades dos colegas e a visualização do *feedback* recebido das avaliações.

Após a realização do experimento, seguiu-se as orientações de [Coutinho, 2011] para a coleta de dados, utilizando-se de técnicas e instrumentos diversificados, sendo: questionários aplicados com os estudantes e o docente, análise documental de dados gerados durante a realização do experimento e reuniões com o docente da disciplina. Os resultados e as discussões do experimento são apresentados no capítulo seguinte.

4. Resultados e discussões

Dos 63 participantes, 31 participaram do grupo experimental e 32 do grupo de controle. No grupo de controle 87% eram do gênero masculino e 13% do gênero feminino, sendo a faixa etária de 26 a 35 anos a com maior quantidade de estudantes (41%). No grupo experimental 93% eram do gênero masculino e 7% eram do gênero feminino, sendo a faixa etária de 16 a 25 anos a com maior quantidade de estudantes (80%). Percebe-se que, em ambos os grupos, o gênero masculino era a grande maioria e que o grupo experimental tinha uma média de idade menor do que o grupo de controle.

Em relação ao módulo do professor, foram aplicados os cálculos da amplitude interquartil para detectar possíveis notas fora do padrão. A Figura 3 mostra algumas

situações em que ocorreu essa detecção, sendo que nesta ocasião os alunos 2, 4, 8 e 14 (primeira coluna) do grupo de controle receberam notas (quatro colunas seguintes) que fugiam do padrão. Destaca-se que no grupo de controle 11 alunos tiveram uma nota considerada *outlier* e no grupo experimental foram 5 alunos na mesma situação.

| Par Avaliado | Avaliador 1 | Avaliador 2 | Avaliador 3 | Avaliador 4 | Média Final | Desvio Padrão |
|--------------|-------------|-------------|-------------|-------------|-------------|---------------|
| 2 | 65 | 80 | 80 | 90 | 78,75 | 10,30 |
| 4 | 70 | 50 | 45 | 50 | 53,75 | 11,08 |
| 8 | 85 | 25 | 25 | 35 | 42,5 | 28,72 |
| 14 | 75 | 100 | 75 | 75 | 81,25 | 12,5 |

Figura 3 - Notas outliers.

A partir do questionário respondido pelo docente da disciplina e reunião posterior com o mesmo, a detecção das notas *outliers* foi útil para que o professor pudesse realizar uma análise das avaliações e descartar possíveis notas que considerava necessário. A avaliação do módulo do professor ocorreu com apenas um docente, pois era necessário que o professor acompanhasse todo o processo do experimento, sendo necessário realizar novos experimentos e análises para verificar o real impacto do módulo do professor para o processo da OPA.

Em relação ao módulo do estudante, para responder às perguntas de pesquisa formuladas, um questionário com oito perguntas foi apresentado aos estudantes, sendo seis em escala likert (com escalas indo de discordo totalmente até concordo totalmente) e duas questões abertas. Destaca-se que os resultados a seguir, que fazem relação à PP1 e a PP2, foram apresentados e debatidos anteriormente em [Oliveira et al., 2019b].

Em relação a PP1, foi possível identificar que os estudantes do grupo experimental sentiram maior conforto com a maneira que os pares eram formados em detrimento ao grupo de controle, sendo que no grupo experimental 75% concordaram parcial ou totalmente e no grupo de controle 69% concordaram parcial ou totalmente ao serem perguntados se eles estavam confortáveis com a formação dos seus pares avaliadores.

Outras perguntas faziam referência ao conforto em avaliarem e serem avaliados, especificamente. Os estudantes do grupo experimental concordaram parcial ou totalmente em 81% dos casos que se sentiam confortáveis com os colegas que avaliaram e 72% do grupo de controle responderam o mesmo. Em relação aos colegas que lhes avaliaram, o grupo de controle e experimental obtiveram porcentagem semelhante, sendo 69% e 68%, respectivamente, os estudantes que se sentiram confortáveis.

A partir destes resultados, que demonstram haver influência da formação dos pares heterogêneos, utilizando o AFP-OPA, para trazer maior conforto do estudante no processo da OPA, novas avaliações serão realizadas para compreender os indícios iniciais encontrados. Leva-se em consideração que, mesmo na pergunta em que as porcentagens foram bem semelhantes, os alunos que discordaram parcial ou totalmente que sentiram-se confortáveis foram 12% no grupo de controle contra apenas 6% no grupo experimental.

Para responder a PP2, foi perguntado aos estudantes se eles consideravam relevantes os apontamentos recebidos pelos seus pares avaliadores, ou seja, o *feedback* de sua atividade. Nesta situação, o grupo de controle concordou parcial ou totalmente em maior número (65%) que o grupo experimental (48%). Outra pergunta feita foi em relação as diferentes perspectivas que os estudantes receberam em suas avaliações, se houve múltiplas contribuições para melhorar seu trabalho. Novamente o grupo de controle obteve melhores resultados, sendo 79% dos estudantes que concordavam que tinham recebido diferentes considerações de seus avaliadores frente a 75% do grupo experimental.

Buscando compreender as razões para que o grupo de controle tenha alcançado melhores resultados no experimento com relação a PP2, foi verificado que, por razão do software “OPA!” utilizar primariamente rubricas para a avaliação e os comentários escritos serem opcionais, apenas 4,8% dos estudantes utilizaram a caixa de comentários para realizar a avaliação no grupo experimental e 6,4% no grupo de controle. Ou seja, o *feedback* recebido pelos estudantes foi majoritariamente quantitativo e não qualitativo, o que pode ter dificultado aos estudantes responderem tais perguntas. Conforme citam [Xiao e Lucking, 2008], há uma melhora nas habilidades dos estudantes que recebiam o *feedback* qualitativo além dos dados numéricos nas avaliações, indicando que o pouco *feedback* de comentários recebidos pode ter dificultado o experimento.

Foi perguntado também, em questão aberta, no que a OPA contribuía para o aprendizado dos estudantes. A Figura 4 traz uma nuvem de palavras, com a mineração dos termos mais recorrentes a partir das respostas dos estudantes. Termos como conhecimento, aprendizado, melhorar associados com diferentes e avaliar mostram que parte dos estudantes consideraram que o ganho de aprendizagem estava na verdade na realização das avaliações de seus colegas e não no *feedback* que recebiam de suas atividades. A partir da leitura dos relatos, percebeu-se que os estudantes consideraram que avaliar os trabalhos dos colegas traziam a eles diferentes pontos de vista para melhorarem seus próprios trabalhos.



Figura 4 - Nuvem de palavras.

Entre as dificuldades relatadas pelos estudantes na realização da atividade está a dificuldade de acesso à internet nos laboratórios que foram realizados os experimentos. Um estudante citou que teve dificuldade em utilizar as rubricas para a avaliação por elas não apresentarem exemplos prévios sobre o que era certo ou não para realizar as avaliações. O docente da disciplina pontuou a dificuldade com o acesso a internet como o principal problema encontrado para realizar as atividades, mas sem maiores dificuldades no andamento do experimento.

5. Considerações finais

A partir da pesquisa realizada, foi possível coletar dados da relação entre a formação dos pares e o desconforto que estudantes sentem ao realizarem o processo da OPA, pouco estudado na literatura da área de informática na educação. O objetivo proposto inicialmente foi o de desenvolver uma ferramenta educacional para diminuir o desconforto dos estudantes e melhorar sua participação na OPA.

Conforme debatido, os dados do experimento mostraram que os participantes do grupo experimental, em relação ao grupo de controle, sentiram-se mais confortáveis com a forma que seus pares foram formados, com quem eles avaliaram e por quem eles foram avaliados. Já em relação ao aproveitamento do *feedback* e a diversidade entre as avaliações recebidas, o resultado foi o oposto. Há a possibilidade de que a não obrigatoriedade de fazer comentários escritos tenha diminuído a diversidade das avaliações e afetado nos resultados obtidos. Também destaca-se que o ganho de aprendizado dos estudantes não foi no *feedback* recebido de suas avaliações a partir das rubricas, mas sim das avaliações que realizaram das atividades de seus colegas.

Estudos como o de [Sitthiworachart e Joy, 2008] apresentam os relatos de estudantes em relação ao desconforto mas não debatem e nem o associam diretamente a uma ou mais etapas da OPA. Por esse e outros motivos, como a diversidade de metodologias empregadas, há uma dificuldade em comparar os resultados obtidos na pesquisa com outros trabalhos similares. Também são necessários novos estudos para compreender melhor esta relação a partir dos indícios iniciais encontrados, assim como o estudo de outras possíveis dificuldades.

Surgem, portanto, diversos trabalhos futuros: aplicação de outras características e atributos dos estudantes para a formação dos pares; aplicação de novas técnicas computacionais além do *k-means*; utilização de sistemas tutores inteligentes para acompanhar o estudante durante a realização das atividades e avaliações para mitigar os problemas encontrados e debatidos na seção 4. Entende-se que este estudo contribui com a área de Informática na Educação, em especial na avaliação da aprendizagem com suporte computacional, a partir da ferramenta educacional desenvolvida e também do experimento realizado. Foi dada atenção para pesquisas de *Online Peer Assessment* como forma de avançar os estudos ligados a estas, contribuindo para os processos de avaliação que se dão a partir de ferramentas online.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referências

- Alves, J., Silva, W., Brito, J. O., e Oliveira, E. (2018) Avaliação em Pares e Autoavaliação: Um Modelo Estatístico Para Perfilação de Alunos. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE), p. 1653-1662.
- Boud, D.; Cohen, R. e Sampson, J. (2014) Peer learning in higher education: Learning from and with each other. Routledge.
- Bouzidi, L. e Jaillet, A. (2009) Can online peer assessment be trusted?. Journal of Educational Technology & Society, v. 12, n. 4, p. 257, 2009.
- Briggs-Myers, I.; Hammer, A.; McCauley, M. e Quenk, N. (2003) MBTI Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator, 3rd. Palo Alto, CA: Consulting Psychologists Press.
- Chandola, V.; Banerjee, A. e Kumar, V. (2007) Outlier detection: A survey. ACM Computing Surveys, v. 14, p. 15.
- Chen, C. (2010) The implementation and evaluation of a mobile self-and peer-assessment system. Computers & Education, v. 55, n. 1, p. 229-236.
- Coutinho, C. (2011) Metodologias de investigação em ciências humanas. Coimbra: Almedina.
- Couto, G.; Bartholomeu, D. e Montiel, J. M. (2016) Estrutura interna do Myers Briggs Type Indicator (MBTI): evidência de validade. Avaliação Psicológica, v. 15, n. 1, p. 41-48.
- Fernandes, D. (2005) Avaliação alternativa: perspectivas teóricas e práticas de apoio. Livro do 3.º Congresso Internacional Sobre Avaliação na Educação. Futuro Eventos, p. 79-92.
- Keirse, D. e Bates, M. (1998) Please understand me II. Del Mar: Prometheus Nemesis Book Company.
- Oliveira, L. C. (2019) Arcabouço para a formação de pares de estudantes para Online Peer Assessment. Dissertação de Mestrado, Universidade Federal do Paraná, Brasil.
- Oliveira, L.; Rosa, S. S. e Pimentel, A. (2019a) Revisão Sistemática da Literatura: Formação de Grupos na Aprendizagem Colaborativa com Suporte Computacional. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). p. 1955.
- Oliveira, L.; Oliveira, W. C.; Rosa, S. S. e Pimentel, A. (2019b) Arcabouço de Formação de Pares para Online Peer Assessment. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). p. 1945.
- Pereira, A.; Oliveira, I. e Tinoca, L. (2010) A Cultura de avaliação: que dimensões? I Encontro Internacional TIC e Educação.

- Rosa, S. S.; Coutinho, C. P. e Flores, M. A. (2017) Online Peer Assessment no ensino superior: uma revisão sistemática da literatura em práticas educacionais. *Avaliação: Revista da Avaliação da Educação Superior*, v. 22, n. 1, p. 55-83.
- Sitthiworachart, J. e Joy, M. (2008) Computer support of effective peer assessment in an undergraduate programming class. *Journal of computer assisted learning*, v. 24, n. 3, p. 217-231.
- Stahl, G.; Koschmann, T. e Suthers, D. (2006) Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge, UK: Cambridge University Press.
- Topping, K. J. (2009) Peer assessment. *Theory into practice*, v. 48, n. 1, p. 20-27.
- Wen, M. L. e Tsai, C. (2008) Online peer assessment in an inservice science and mathematics teacher education course. *Teaching in Higher Education*, v. 13, n. 1, p. 55-67.
- Xiao, Y. e Lucking, R. (2008) The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. *The Internet and Higher Education*, v. 11, n. 3-4, p. 186-193.