

USANDO AVALIAÇÃO POR PARES PARA ENCORAJAR A AUTOAVALIAÇÃO

Igor dos Santos Montagner– igorsm1@insper.edu.br
INSPER - Instituto de Pesquisa e Ensino
Rua Quatá, 300
04546-042 – São Paulo – SP

Marcelo Hashimoto– marcelo.hashimoto@insper.edu.br
INSPER - Instituto de Pesquisa e Ensino
Rua Quatá, 300
04546-042 – São Paulo – SP

Fábio Miranda– fabiomiranda@insper.edu.br
INSPER - Instituto de Pesquisa e Ensino
Rua Quatá, 300
04546-042 – São Paulo – SP

Resumo: A avaliação por pares pode ser vantajosa tanto por fatores pedagógicos quanto operacionais. É uma prática que treina os estudantes em dar feedback, habilidade que útil para seu futuro profissional, e, ao mesmo tempo, permite devolutivas mais ágeis, contornando o gargalo representado pela disponibilidade de tempo do instrutor. A avaliação por pares permite, ainda, que o aluno reflita sobre seu próprio nível de aprendizado ao avaliar a produção de um colega. Este trabalho relata uma experiência sistemática de avaliação por pares, amparada por instruções detalhadas de atribuição de notas, que foi realizada em uma disciplina do 5o. semestre de um curso de Engenharia de Computação. Um questionário aferiu a percepção dos estudantes sobre a atividade. São apresentados os resultados dos questionários e uma reflexão sobre as respostas dos alunos.

Palavras-chave: Avaliação entre pares. Rubricas. Educação em Engenharia

1 INTRODUÇÃO

O processo de avaliação por pares propõe que existem vantagens pedagógicas e logísticas em criar atividades nas quais os próprios estudantes atribuem notas aos trabalhos de colegas (Sadler, 2006). Do ponto de vista pedagógico, refletir sobre diferentes soluções para uma questão pode ser suficiente para desenvolver os objetivos de aprendizado (Sadler, 2006). Além disso, é necessário aplicar processos cognitivos de ordem superior para atribuir notas (Bloom, 1971). Do ponto de vista logístico, a avaliação por pares elimina o principal gargalo na avaliação: o tempo do professor. Ao permitir que estudantes atribuam notas, diminuimos significativamente o tempo de espera por feedback (Boud, 1989).

Além de aspectos puramente acadêmicos, podemos argumentar também que a avaliação por pares (e também a autoavaliação) fará parte do futuro profissional de muitos estudantes e,

portanto, exercitar esta habilidade durante a graduação torna sua preparação para o futuro mais completa.

Uma das preocupações frequentes em relação a esse processo é a possibilidade de imprecisão das notas atribuídas por estudantes, mas diversos estudos mostram que elas podem ser próximas das atribuídas por professores. Uma meta análise de 48 estudos realizada por Falchikov (2000) mostrou que as notas atribuídas por estudantes tendem a seguir de forma clara as notas atribuídas pelo instrutor quando a atribuição de notas é amparada por critérios bem compreendidos e quando resultados acadêmicos, em vez de profissionais, são medidos. Resultados similares são reportados em (Sadler, 2006).

Neste trabalho, descrevemos uma atividade de avaliação por pares executada em uma turma do 5o semestre de Engenharia de Computação. Essa atividade teve como objetivos principais melhorar a autoavaliação dos estudantes e entender melhor os critérios de desempenho e qualidade adotados na disciplina. Após um quiz de uma hora, cada estudante recebeu as soluções junto com uma página de instruções de correção e iniciou-se a atividade de avaliação por pares.

Os trabalhos (Cho *et al*, 2006) e (Hafner e Hafner, 2003) dão ênfase, além da atividade de avaliação por pares, também nas instruções de correção, citando o uso de rubricas como fator importante para resultados positivos em relação à acurácia das notas produzidas. Diversos benefícios da utilização de rubricas também são descritos em (Reddy e Andrade, 2010). As instruções de correção apresentadas neste trabalho não seguem o formato tradicionalmente atribuído a rubricas, porém argumentamos que servem o mesmo propósito: coletar evidências de que um certo objetivo de aprendizado foi alcançado. Portanto, é possível que alguns benefícios do uso de rubricas também sejam observados nesta atividade.

Para quantificar a percepção dos estudantes sobre a atividade foi criado um questionário cujo objetivo era medir tanto o entendimento das instruções de correção quanto entender se a atribuição de nota ao trabalho de um colega gerou reflexão sobre o próprio desempenho e se a atividade de avaliação por pares causou sentimentos negativos entre os estudantes.

Na Seção 2, descrevemos a atividade de avaliação entre pares e apresentamos exemplos de questões usadas no quiz e do material de atribuição de nota fornecido aos estudantes. Na Seção 2.1, discutimos a relação entre as instruções de correção criadas e rubricas tradicionais. Na Seção 3, analisamos os resultados da atividade e julgamos se foi bem-sucedida. Na Seção 3.1, discutimos a execução da atividade em sala de aula e enumeramos o que funcionou e o que não funcionou como planejado. Na Seção 4, chegamos a algumas conclusões e discutimos futuros rumos para o trabalho.

2 DESCRIÇÃO DA ATIVIDADE PROPOSTA E DAS INSTRUÇÕES DE ATRIBUIÇÃO DE NOTAS

A atividade de avaliação por pares proposta é composta por duas partes, em que cada uma dura uma hora. Na primeira parte, estudantes resolvem um quiz contendo 3 questões de baixa a média complexidade. Na segunda parte, cada estudante recebe o quiz de um colega e um conjunto detalhado de instruções de correção, acompanhado das soluções de cada exercício. Foi avisado que mais de uma solução era possível e que eles deveriam interpretar as soluções com cuidado.

Esta atividade foi executada no contexto de uma disciplina oferecida ao 5o semestre da Engenharia de Computação, que combina arquitetura de computadores e conceitos de sistemas operacionais. Os seguintes objetivos de aprendizagem foram medidos no quiz:

1. o estudante é capaz de identificar expressões aritméticas em Assembly x64 e traduzi-las para código C limpo e legível;
2. o estudante é capaz de identificar estruturas de controle, como condicionais e laços, em Assembly x64 e transformá-las em código C estruturado;
3. o estudante é capaz de identificar expressões booleanas em Assembly x64 e transformá-las em código C legível.

Um exemplo de questão e instrução de correção é mostrado abaixo. A instrução de correção inclui uma possível solução, quais elementos devem ser identificado na questão e quanto cada um deles vale. Cada elemento se relaciona com um dos objetivos de aprendizagem apresentados acima.

Quadro 1 - Exemplo de questão e sua instrução de correção.

<p>Traduza a função abaixo para C linha a linha usando somente <code>if-goto</code> e <code>labels</code>.</p> <pre> 00000000000005fa <ex1>: 5fa: test %edx,%edx 5fc: jne 607 5fe: lea (%rsi,%rsi,4),%rdx 602: lea (%rdx,%rdi,1),%rax 606: retq 607: mov %rdi,%rax 60a: sub %rsi,%rax 60d: add %rax,%rax 610: retq </pre>	<p><u>Solução:</u></p> <pre> long ex1(long a, long b, int flag){ long retval; long temp; if (flag != 0) goto LB1; temp = b + 4 * b; retval = a + temp; return retval; LB1: retval = a; retval -= b; retval += retval; return retval; } </pre> <p><u>Instrução de correção:</u> cada item abaixo vale 0.25;</p> <ul style="list-style-type: none"> • A expressão booleana <code>flag != 0</code> foi identificada corretamente • O fluxo de controle do código não foi modificado: <code>jumps</code>, <code>labels</code> e <code>return</code> estão posicionados corretamente. • O bloco <code>flag==0</code> computa <code>a + 5b</code> • O bloco <code>flag!=0</code> computa <code>2a + 2b</code> <p>Se o erro abaixo for identificado, diminua -</p>
---	--

	<p>0.125:</p> <ul style="list-style-type: none">• O código usa a variável <code>flag</code> para os cálculos. Uma nova variável do tipo <code>long</code> é necessária. Caso contrário todos os cálculos são convertidos implicitamente para <code>int</code>.
--	--

Após resolver o quiz, cada estudante recebeu o quiz de um colega e foi instruído a corrigir cada questão usando os critérios apresentados nas instruções de correção e de acordo com seu entendimento das soluções. O professor estava disponível para responder perguntas sobre os critérios e, se pedido explicitamente, para ajudar estudantes a entenderem soluções não usuais. Porém, o professor não interferiu nas correções feitas pelos estudantes.

Foi avisado que as correções seriam revisadas pelo professor e que seria necessário se identificar. Isto foi feito por duas razões. Por um lado, gostaríamos que os estudantes fossem responsáveis pela qualidade de suas correções e exigir a identificação pode ajudar a evitar que alguns estudantes sejam levianos na correção. Por outro lado, queríamos evitar um sentimento de responsabilidade e ansiedade pelo sucesso ou fracasso de seus colegas. Por conta disso foi reforçado, antes e durante a atividade, que as notas seriam revisadas e que a nota que eles atribuíssem não prejudicaria seus colegas.

2.1 DISCUSSÃO DAS INSTRUÇÕES DE CORREÇÃO

Rubricas frequentemente são descritas como compostas de duas características: um conjunto de critérios para avaliação e níveis definidos de qualidade ou desempenho para cada critério (Brookhart, 2013; Andrade, 2000). O critério deve estar relacionado com os objetivos de aprendizado para cada atividade e os níveis de desempenho deveriam refletir o domínio daquele objetivo de aprendizado específico. As instruções de atribuição de nota mostradas no Quadro 1 não se encaixam de forma estrita nesta definição. Entretanto, servem para a mesma finalidade: coletar evidência de que determinado objetivo de aprendizado foi alcançado. Cada item nas instruções de correção está diretamente relacionado com um único objetivo de aprendizado (descrito acima do Quadro 1) e indica se uma solução apresenta evidência de domínio daquele objetivo de aprendizado.

Por exemplo, o critério "o bloco `flag==0` calcula a expressão $a + 5b$ " e "o bloco `flag!=0` computa $2a + 2b$ " afere se o estudante é competente no objetivo de aprendizado 1. Uma vez que cada objetivo de aprendizado é avaliado em mais de uma questão, os resultados agregados indicam o desempenho do estudante em relação a todos os objetivos de aprendizado.

Outro ponto importante durante a criação das instruções de correção foi a necessidade de torná-las de fácil entendimento para os alunos. Em (Andrade e Du, 2005) é reportado que uma porção significativa dos estudantes não leem a rubrica em sua totalidade. Optamos, então, por um formato menos tradicional, mas que facilita a correção pelo estudante ao relacionar cada item da correção com um elemento mostrado na solução. Além disso, podemos argumentar que, ao quebrar a solução em várias pequenas partes, facilitamos seu entendimento. O mesmo exemplo acima poderia ser codificado em uma rubrica com o critério “Aritmética inteira em

x64 e sua tradução para código C” e níveis de desempenho descritos como “Básico - o estudante identifica expressões simples quem usem somente uma instrução aritmética simples”. Esta descrição, porém, não ajuda o estudante-corretor a identificar na solução analisada se este objetivo foi alcançado ou não. Instruções de correção que priorizem o entendimento pelo aluno tornam o processo de avaliação por pares mais transparente e podem melhorar a acurácia das notas.

3 RESULTADOS

Nesta seção, discutimos os resultados da atividade e analisamos a percepção dos estudantes de maneira qualitativa. Também discutimos o planejamento original da atividade e como ele foi modificado durante sua execução.

Após a execução da atividade, os alunos responderam um questionário com o objetivo de mensurar sua percepção da atividade. As 3 dimensões abaixo foram foco do questionário.

1. A atividade ajuda o aluno a avaliar seu próprio progresso (autoavaliação)?
2. As instruções de correção estão claras (entendimento dos critérios)?
3. A atividade causa sentimentos negativos ou animosidade entre estudantes (implicações sociais e pessoais)?

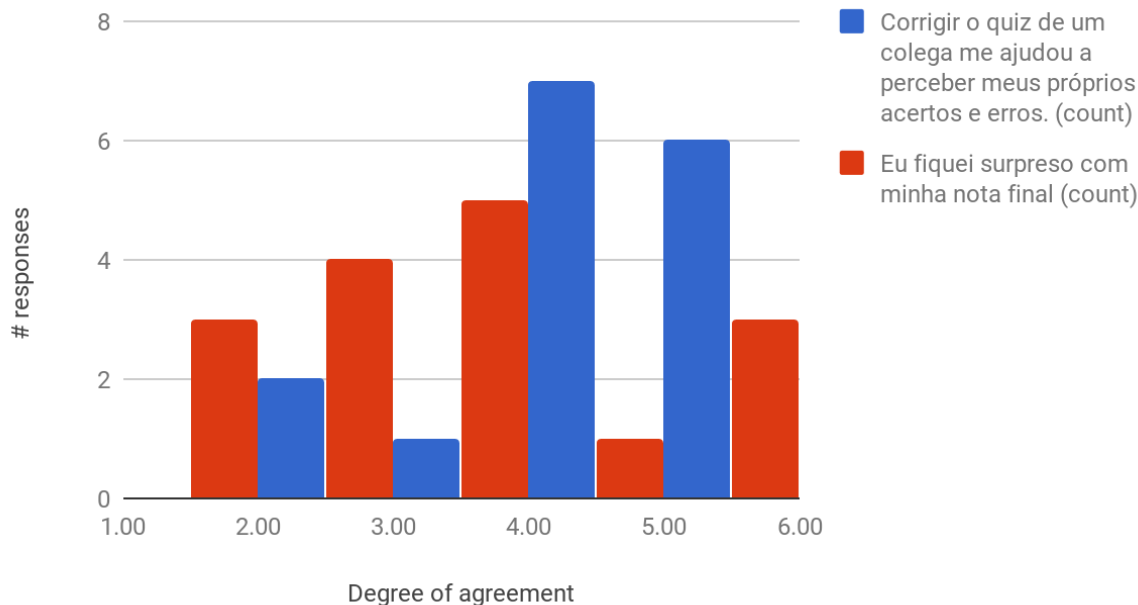
Foram apresentadas as seguintes 5 afirmações e foi pedido que cada estudante avalie seu grau de concordância com cada uma, sendo que 1 significa "Discordo totalmente" e 5 "Concordo totalmente". Recebemos 16 respostas de um total de 20 estudantes.

1. Corrigir o quiz de um colega me ajudou a perceber meus próprios acertos e erros.
2. Eu me senti confortável corrigindo o quiz de um colega.
3. Após a correção eu estava confiante da nota que atribuí.
4. Eu gostaria de repetir esta atividade.
5. Eu fiquei surpreso com minha nota final.

As afirmações 1 (em azul) e 5 (em vermelho) são relacionadas com a dimensão Autoavaliação. O histograma das respostas é apresentado na Figura 1.

Figura 1 -Histograma do grau de concordância para afirmações relacionadas a Auto-avaliação

Auto-avaliação



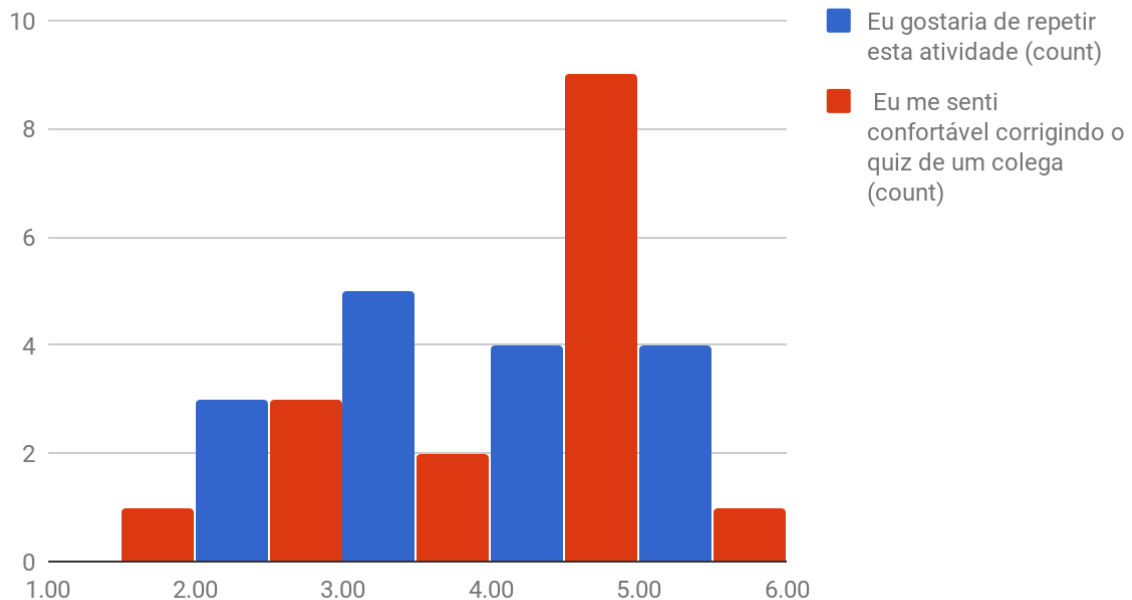
A maioria dos estudantes (13 de 16) concorda, ao menos parcialmente, que a atividade os ajudou a entender seus erros e acertos. Além disso, nenhum estudante discordou completamente desta afirmação. Isto indica que, do ponto de vista dos estudantes, a atividade serviu como autoavaliação apesar deles estarem corrigindo o trabalho de um colega.

A quinta afirmação ("Eu fiquei surpreso com minha nota final") teve um padrão de respostas mais uniforme. Apesar dos resultados parecerem viesados em direção à discordância, não podemos concluir muito a partir destas respostas, pois durante e após a aplicação do questionário houve dúvidas sobre esta afirmação. Existem duas interpretações possíveis e igualmente válidas: um estudante pode estar surpreso pois esperava ir mal no quiz ou ele pode estar surpreso pois a atividade não o ajudou a estimar sua nota de maneira satisfatória. Apresentamos estes resultados por completude.

As afirmações 2 (em vermelho) e 4 (em azul) são relacionadas aos impactos sociais e pessoais. O histograma das respostas é mostrado na Figura 2.

Figura 2 - Histograma do grau de concordância das afirmações relativas a impactos sociais e pessoais.

Impactos sociais e pessoais



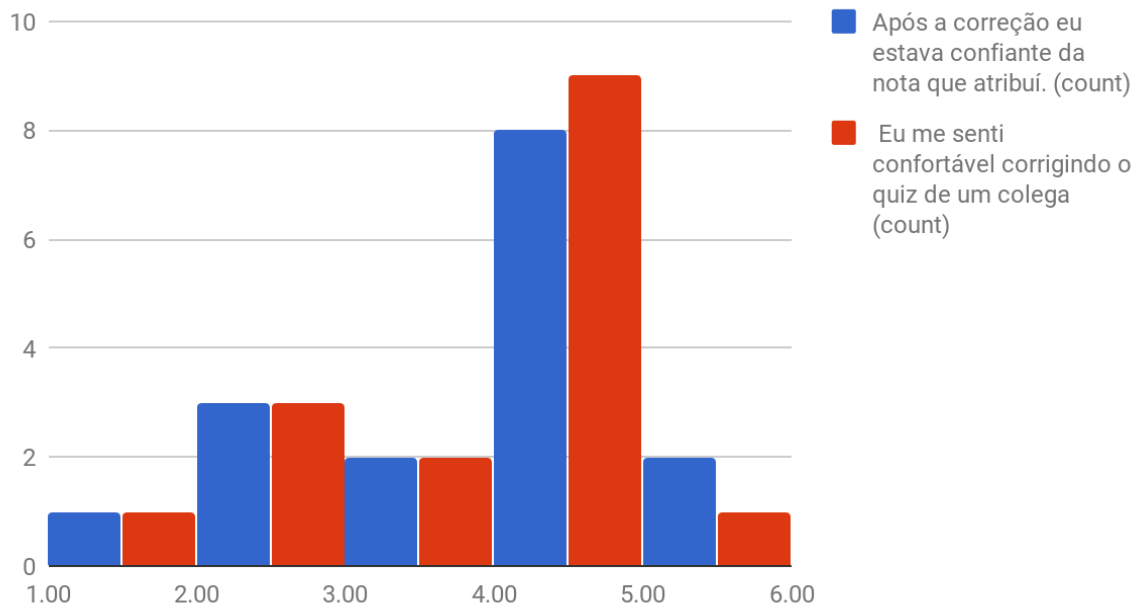
A afirmação 4 (em azul) possui resultados mistos. Apesar de nenhum estudante ser completamente contra sua repetição, a moda do histograma é "Neutro". Como existem mais concordância que discordância concluímos que a repetição da atividade não seria má recebida.

O padrão de concordância da afirmação 2 (em vermelho) mostra que, para a maioria dos estudantes, (12 de 16) a atividade não causou sentimentos negativos. Porém, precisamos levar em conta que 4 estudantes sentiram algum desconforto. A próxima análise ajuda a entender melhor estas respostas.

A afirmação 3 (em azul) está relacionada ao entendimento das instruções de correção. Também analisamos neste item as respostas da afirmação 2 (em vermelho) pois elas também estão relacionadas a esta dimensão. O histograma das respostas é mostrado na Figura 3.

Figura 3 - Histograma do grau de concordância para afirmações relacionadas ao entendimento das instruções de correção.

Entendimento das instruções de correção



O padrão de respostas das afirmações 2 (em vermelho) e 4 (em azul) é quase idêntico. Isso pode explicar a razão do desconforto alguns estudantes: eles não estavam confiantes em suas correções. Por outro lado, novamente há uma concordância relativamente forte, com uma porção significativa dos estudantes (10 de 16) sentindo-se confiante em relação à nota atribuída. Dado que esta foi a primeira experiência de correção da turma, estes resultados podem ser considerados satisfatórios.

Os padrões de resposta registrados indicam que a atividade cumpriu seus objetivos. Foi encorajada autoavaliação e não parece ter havido sentimentos negativos ligados à sua execução. Além disso, existem indicadores de que grande parte dos alunos compreendeu os critérios de correção, tornando o processo de avaliação mais transparente.

3.1 DISCUSSÃO SOBRE A EXECUÇÃO DA ATIVIDADE

Apesar de termos obtidos resultados satisfatórios, a execução da atividade não seguiu o plano original. Pretendíamos corrigir as questões em sequência para permitir fazer uma revisão do conteúdo durante a correção. Porém, cada estudante seguiu um ritmo completamente diferente e em pouco tempo cada um estava corrigindo uma questão diferente. A partir de então, desistimos da sincronização da correção entre estudantes e passamos para um modelo assíncrono em que o professor circulava pela sala tirando dúvidas sobre os critérios de correção. Para repor este momento de revisão, foi elaborada uma atividade prática no começo da aula seguinte, revisando os principais erros no quiz. Estes erros não foram repetidos pelos alunos na avaliação intermediária que ocorreu 2 semanas após a atividade proposta.

Nos outros aspectos a atividade ocorreu sem grandes problemas. O tempo disponibilizado foi suficiente para todos os estudantes e eles, aparentemente, fizeram as correções com seriedade mesmo sabendo que as notas seriam revisadas. Além disso, as notas atribuídas pelos alunos foram bastante próximas das notas atribuídas pelo professor, corroborando a sensação de confiança que eles reportaram nos questionários.

4 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho, relatamos a experiência de uma atividade de avaliação por pares feita com alunos do 5o semestre de Engenharia de Computação. Cada aluno corrigiu a atividade de um colega baseado em um conjunto de instruções de correção e uma análise qualitativa dos dados indica que a atividade encorajou uma reflexão auto avaliativa em relação ao desempenho no quiz. Além disso, alunos não reportaram de maneira significativa sentimentos negativos associados à execução da atividade e não receberiam mal uma reedição da atividade descrita. Os resultados obtidos foram encorajadores e as notas produzidas na atividade foram bastante próximas às notas atribuídas pelo professor.

No futuro, pretendemos fazer uma análise quantitativa dos dados comparando, as notas atribuídas por alunos às atribuídas pelo professor e comparando as notas do quiz às notas da prova intermediária. Além disso, gostaríamos de executar esta atividade com um número maior de alunos, se possível de turmas e disciplinas diferentes.

Agradecimentos

Agradecemos ao Insper pelo apoio durante a realização deste trabalho.

REFERÊNCIAS

ANDRADE, Heidi. **Using rubrics to promote thinking and learning**. Educational leadership. 2000

ANDRADE, Heidi; DU Ying., **Student Perspectives on Rubric-Referenced Assessment** Educational & Counseling Psychology Faculty Scholarship 2; 2005.

BLOOM, Benjamin. **Taxonomy of Educational Objectives: The Classification of Educational Goals: By a Committee of College and University Examiners**. David McKay; 1971.

BROOKHART, Susan. **How to create and use rubrics for formative assessment and grading**. Ascd, 2013

CHO, Kwangsu; SCHUNN, Christian; WILSON, Roy. **Validity and Reliability of Scaffolded Peer Assessment of Writing From Instructor and Student Perspectives**. Journal of Educational Psychology. 98. 2006

FALCHIKOV, Nancy, GOLDFINCH, Judy. **Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks**. Review of educational research. 2000

HAFNER, John; HAFNER, Patti **Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating**, International Journal of Science Education, 25:12, 2010

REDDY, Y. Mallini; ANDRADE, Heidi. **A review of rubric use in higher education**. Assessment & evaluation in higher education. 2010 Jul 1;35(4):435-48.

SADLER, Philip; GOOD, Eddie. **The impact of self-and peer-grading on student learning**. Educational assessment. 2006

USING PEER EVALUATION TO PROMOTE SELF-ASSESSMENT

Using peer evaluation to promote self-assessment

***Abstract:** Peer evaluation presents advantages from both pedagogical and an operational perspective. It is a process that trains students in providing feedback, a useful skill for their future professional lives, and, at the same time, allows faster feedback, bypassing the bottleneck represented by the instructor's availability. Peer evaluation also allows the student to reflect about his or her own level of learning when assessing the work of a colleague. This work reports a systematic peer evaluation experience, supported by detailed grading instructions, that was carried out in a 5th semester Computer Engineering course. A survey was used to assess the students' perception about the activity. We present the results of the survey and a reflection about the students' responses to it.*

Keywords: Peer Assessment. Rubrics. Engineering Education