



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

FERNANDA DA SILVA

**UM CARÁTER DE VALIDADE DA AVALIAÇÃO NA  
PERSPECTIVA DA EDUCAÇÃO MATEMÁTICA  
REALÍSTICA**

---

Londrina  
2018

FERNANDA DA SILVA

**UM CARÁTER DE VALIDADE DA AVALIAÇÃO NA  
PERSPECTIVA DA EDUCAÇÃO MATEMÁTICA  
REALÍSTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ensino de Ciências e Educação Matemática da Universidade Estadual de Londrina como requisito parcial à obtenção do título de Mestre.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Regina Luzia Corio de Buriasco.

Lodrina  
2018

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

Silva, Fernanda da .

Um caráter de validade da avaliação na perspectiva da Educação Matemática Realística / Fernanda da Silva. - Londrina, 2018.

46 f. : il.

Orientador: Regina Luzia Corio de Buriasco.

Dissertação (Mestrado em Ensino de Ciências e Educação Matemática) - Universidade Estadual de Londrina, Centro de Ciências Exatas, , 2018.

Inclui bibliografia.

1. Educação Matemática Realística - Tese. 2. Avaliação didática - Tese. 3. Validade - Tese. I. Corio de Buriasco, Regina Luzia . II. Universidade Estadual de Londrina. Centro de Ciências Exatas. . III. Título.

FERNANDA DA SILVA

**UM CARÁTER DE VALIDADE DA AVALIAÇÃO NA  
PERSPECTIVA DA EDUCAÇÃO MATEMÁTICA REALÍSTICA**

Dissertação apresentada ao Programa de Pós-Graduação em Ensino de Ciências e Educação Matemática da Universidade Estadual de Londrina como requisito parcial à obtenção do título de Mestre.

**BANCA EXAMINADORA**

---

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Regina Luzia  
Corio de Buriasco  
Universidade Estadual de Londrina - UEL

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Maria do Socorro Taurino  
Universidade Nove de Julho - UNINOVE

---

Prof. DR. Jorge Luiz Vargas Prudencio de  
Barros Pires  
Universidade Estadual de Londrina –  
UNESP

Londrina, 05 de março de 2018.

*Dedico este trabalho a todos os  
alunos com quem aprendi algo para  
minha construção como ser humano.*

## **AGRADECIMENTOS**

- À minha mãe, Maria, por ser a pessoa com quem eu posso contar sempre. Pernambucana “cabra da peste”, simples e prestativa. Sou muito grata pelo seu amor, pelo incentivo, por suas orações, pelo apoio financeiro e principalmente por tudo que com ela aprendo.

- À minha orientadora, Regina, por acreditar em mim e me incentivar, pelo auxílio no desenvolvimento deste trabalho, por todas as oportunidades de aprendizagem que me ofereceu, desde a graduação, sobretudo por aquelas que me fizeram refletir a respeito de minha prática docente.

- Aos membros do GEPEMA, pelas contribuições a este trabalho e também pelos momentos de aprendizagem. Em especial, Ana Cláudia e José Emídio que, além disso, estiveram comigo sempre que precisei.

- Aos membros da banca, Prof.<sup>a</sup>. Dr.<sup>a</sup>. Maria do Socorro Taurino, Prof.<sup>a</sup>. Dr.<sup>a</sup>. Edilaine Regina dos Santos e Prof. Dr. Jorge Luiz Vargas Prudêncio de Barros Pires por aceitarem o convite e pelas considerações apresentadas.

- Ao Rodrigo, por me ouvir sempre, especialmente nos momentos de aflição e angústia.

- À minha irmã, Fátima, pela tranquilidade que me proporcionou ao sempre fazer companhia à nossa mãe.

- Ao colega de especialização, Jair, pelo incentivo para fazer a seleção e pelas sugestões em relação ao projeto de pesquisa apresentado.

- À colega de mestrado, Karina, pelas conversas, risadas e parceria nas viagens, nestes dois anos.

- À CAPES pela bolsa concedida.

As ideias não são responsáveis  
pelo que os homens fazem delas.  
Werner Heisenberg

SILVA. Fernanda da. **Um caráter de validade da avaliação na perspectiva da Educação Matemática Realística**. 2018. 46 f. Dissertação (Mestrado em Ensino de Ciências e Educação Matemática) – Universidade Estadual de Londrina, Londrina, 2018.

## **RESUMO**

Esta dissertação tem como objetivo apresentar um estudo de características do caráter de validade da avaliação da matemática escolar, na perspectiva da RME, para além dos critérios de validade encontrados na literatura da área, tomando como pressuposto que a escolha de tarefas pode ou não oferecer condições de entendimento e de aprendizagem. Adotada a abordagem qualitativa de cunho especulativo, procura-se neste trabalho identificar um tipo de validade que redimensione o papel da avaliação como prática de investigação e como uma oportunidade de aprendizagem. Para isso, a busca da certeza psicométrica é abandonada, dando lugar a uma abordagem hermenêutica. Uma conclusão é que o caráter de validade na perspectiva da Educação Matemática Realística depende do sucesso das ações desenvolvidas tanto pelo professor quanto pelos alunos nos processos de ensinar e aprender matemática na escola.

**Palavras-chave:** Educação Matemática Realística. Avaliação Didática. Validade.

SILVA. Fernanda da. **A validity character of the assessment from the perspective of Realistic Mathematics Education.** 2018. 46 p. Dissertation (Master's Degree in Mathematics Education and Sciences) – Universidade Estadual de Londrina, Londrina, 2018.

### **ABSTRACT**

This dissertation aims to present a study of characteristics of the validity character of the assessment of school mathematics, from the perspective of Realistic Mathematics Education, in addition to the validity criteria found in the area's literature, assuming that the selection of tasks may or may not offer conditions of understanding and learning. Adopting the speculative qualitative approach, this paper seeks to identify a validity type that re-dimension the role of assessment as an investigative practice and as a learning opportunity. For this, the pursuit for psychometric certainty is abandoned, giving rise to a hermeneutical approach. One conclusion is that the validity character from the perspective of Realistic Mathematics Education depends on the success of the actions developed by both the teacher and the students in the processes of teaching and learning mathematics in the school.

**Keywords:** Realistic Mathematics Education. Didactical Assessment. Validity.

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	10
<b>O ESTUDO E SUA TRAJETÓRIA (O ENLACE)</b> .....	13
<b>DESENLACE</b> .....	26
<b>REFERÊNCIAS</b> .....	34
<b>APÊNDICES</b> .....	39
APÊNDICE A – Resultado da busca da palavra validade em dicionários .....	40
APÊNDICE B – Inventário de fragmentos com alguma referência à validade encontrados nos autores estudados.....	43

## 1 INTRODUÇÃO

No GEPEMA<sup>1</sup>, a avaliação é parte integrante dos processos de ensino e de aprendizagem, como prática de investigação, como mais uma oportunidade de aprendizagem, como uma fonte de formação e informação tanto para os alunos quanto para os professores. Entre outros temas relacionados, os participantes têm se dedicado a estudar avaliação da aprendizagem escolar de matemática na busca de conhecer como alunos de diferentes níveis de escolaridade e professores que ensinam matemática lidam com tarefas de matemática em situação de prova, assim como conhecer a dinâmica da utilização da prova escrita, uma vez que é, talvez, o instrumento mais utilizado em aulas de matemática.

Dando continuidade aos estudos desenvolvidos pelos participantes do GEPEMA, tomando como base a Educação Matemática Realística (RME)<sup>2</sup>, este estudo pretende conhecer características presentes em tarefas de matemática que possam indicar o seu caráter<sup>3</sup> de validade. Por conseguinte, o objetivo deste trabalho reside em apresentar um estudo do caráter<sup>4</sup> de validade da avaliação da matemática escolar, na perspectiva da RME, para além dos critérios de validade encontrados na literatura da área, uma vez que se toma como pressuposto<sup>5</sup> que a escolha de tarefas pode ou não oferecer condições de entendimento e de aprendizagem. Assim, procura-se neste trabalho identificar um tipo de validade que redimensione as aproximações do trabalho desenvolvido na sala de aula com o desenvolvido na avaliação da aprendizagem. Esse

---

<sup>1</sup> GEPEMA – Grupo de Estudos e Pesquisa em Educação Matemática e Avaliação, da Universidade Estadual de Londrina. Outras informações em: <http://www.uel.br/grupo-estudo/gepema/index.html>.

<sup>2</sup> Para fazer referência à Educação Matemática Realística será utilizada a sigla RME da expressão inglesa *Realistic Mathematics Education*.

<sup>3</sup> Baseado em Abbagnano (1982) Caráter tomado aqui como um sinal ou conjunto de sinais, que distingue um objeto e permite reconhecê-lo facilmente entre os outros. O caráter é resultado das opções feitas por um indivíduo, consistindo nas constantes observáveis das suas opções.

<sup>4</sup> Tomado aqui como um conjunto de características próprias.

<sup>5</sup> Tomado aqui como uma conclusão antecipada baseada em indícios e suposições.

redimensionamento busca operacionalizar o papel da avaliação como prática de investigação e como uma oportunidade de aprendizagem.

Tomando como objeto de estudo o caráter de validade da avaliação na matemática escolar, pretende-se:

- inventariar aspectos presentes em textos de autores que escrevem a respeito de avaliação que podem ser tomados como componentes de um caráter de validade.
- Identificar características/elementos que podem compor uma condição<sup>6</sup> de validade ou que estão subjacentes a ela na perspectiva da RME.
- Caracterizar/configurar um caráter de validade na perspectiva da RME.

Neste estudo foi adotada a abordagem qualitativa de cunho especulativo, inspirada em Van der Maren(1996), Clandinin e Connelly (2000); Martineau, Simard e Gauthier, (2001).

Nessa abordagem, a análise é realizada a partir da tematização das informações recolhidas em uma perspectiva de composição de significados. Nessa perspectiva de análise ou interpretação do material, não se buscam critérios para verificar/construir/declarar verdades, uma vez que os autores consideram que, para abordar algo, há mais que três lados.

Para a interpretação, procura-se uma composição de significados como resultante da interação entre o material recolhido e a visão do pesquisador a partir da leitura, escrita e reescrita do material, com isso, fazer escolhas é parte do processo.

Essa abordagem metodológica requer interpretação plausível, construções coerentes, que possibilitem uma sensação contínua de trabalho em desenvolvimento, sempre em processo de tornar-se. A composição de significados é considerada um movimento para frente.

---

<sup>6</sup> Condição: circunstância que determina o caráter ou a existência de um fenômeno, um fato etc.

O processo de análise e interpretação do *corpus* desta investigação foi desenvolvido em torno de três eixos: interpretação, argumentação e recontar. Foram inventariados textos de um mesmo tema e de diferentes autores. A partir da interpretação dos textos inventariados, mediante diferentes comparações, foi feita uma análise interpretativa das relações entre as diferentes ocorrências do caráter de validade nos textos estudados, buscando uma outra configuração plausível (VAN DER MAREN, 1996). Neste trabalho, entende-se “configurar” como “caracterizar”, ressaltando as características do objeto, representando-o com os atributos que o caracterizam em ação.

Este texto compõe-se de três partes. A primeira é esta introdução; a segunda contém o estudo e sua trajetória; a terceira apresenta uma configuração plausível de um caráter de validade na perspectiva da RME.

## O ESTUDO E SUA TRAJETÓRIA (O ENLACE)

Tomando a avaliação da aprendizagem escolar como uma prática de investigação e tendo como premissa que não se avalia em geral, também a validade não é um atributo geral e pode existir em diferentes graus. Assim, a afirmação de que uma prova é válida quando mede aquilo que pretende medir é por demais genérica. Para Vianna (1978, p. 171),

A validade refere-se ao que o teste mede; conseqüentemente, a palavra validade não pode ser usada isoladamente e apresentada abstratamente. Um teste não é válido de modo geral, mas sim em relação a determinado propósito e para um grupo particular com características específicas.

Pensar em configurar um caráter de validade faz emergir duas perguntas iniciais: em que consiste e como pode ser tomado um caráter de validade? Que elementos constituem um caráter de validade (como reconhecer) ou estão subjacentes a ele?

Com o intuito de iniciar esta investigação, buscou-se o significado dos vocábulos em um dicionário de Língua Portuguesa, um Etimológico e um de Filosofia (Apêndice A).

De acordo com o dicionário Houaiss (2009, CD-ROM), o substantivo validade pode ser entendido como a "qualidade ou condição de algo que se encontra em condições de produzir os efeitos dele esperados".

Ainda sobre o significado da palavra validade, Kaplan (1964) afirma que "é o mesmo que o da palavra 'valor': ambos derivam de um termo que significa força (*strength*)" (*apud* MESSICK, 1979, p.7). Em concordância com a afirmação, Cunha (1982) informa que a origem da

palavra válido vem de *adj.* 'orig. sadio, forte' 'ext. que tem valor ou valia' XVI. Do lat. *Validus*

Em Abbagnano (1982, p. 989) encontra-se que um significado para validade diz respeito a "utilidade ou eficiência de um meio ou de um instrumento qualquer", e que na expressão "válido para", o "para" indica a finalidade ou "a função em relação à qual se considera eficiente o instrumento, o meio ou a condição de que se trata."

Como em diferentes épocas a avaliação, assim como os aspectos relacionados a ela, foi entendida de diferentes modos, a validade também assumiu diferentes significados, e diferentes características foram valorizadas ou diferentes tipos de validade. Um inventário de fragmentos contendo alguma referência à validade foi elaborado a partir da leitura de alguns autores escolhidos por conveniência (Apendice B).

De acordo com Vianna (1997), a avaliação inicialmente teve seus fundamentos na psicomетria. Nesse contexto, buscava-se construir instrumentos com o intuito de medir a inteligência, entendendo-a como uma característica inata e imutável no ser humano, determinando a capacidade de aprender de cada um.

A avaliação tradicional<sup>7</sup> se apoiou em ideias que tiveram seu momento na década de 1920, como, por exemplo, a concepção de que a aprendizagem seria linear e sequencial e de que, para ensinar os conhecimentos mais complexos, seria necessário atomizá-los (dividi-los em partes menores)<sup>8</sup>. Ainda hoje existem professores que agem em acordo com essas ideias, ou porque acreditam nelas ou porque não pensam a respeito do assunto, apenas seguem a tradição.

Para a construção dos testes adotava-se um modelo comportamental de aprendizagem. Os resultados eram interpretados com

---

<sup>7</sup> As ideias de avaliação tradicional aqui apresentadas estão contidas em BLACK (2009), MESSICK (1979), MORROW JUNIOR (2013), Moss (1992), VIANNA (1978; 1997a; 1997b; 2002), WILLIAM (1998).

<sup>8</sup> Na perspectiva deste trabalho fazer isso impede a visão do todo.

base em normas, isso significava que os desempenhos dos indivíduos eram comparados com os desempenhos dos demais indivíduos do grupo, de modo a classificá-los. Assim, quanto maior fosse a amplitude de variação dos resultados encontrados melhor a distinção entre eles. Por isso, itens fáceis ou difíceis eram eliminados após um pré-teste. Pouca importância era dada à validade nesse contexto. eram valorizados aspectos técnicos da medida, como a padronização dos resultados, a fidedignidade, entre outros.

A fidedignidade foi tida como uma condição necessária para a validade de um teste. Tradicionalmente, a fidedignidade é associada à reprodutibilidade, ou seja, se um teste for aplicado, um grande número de vezes, a um mesmo grupo, espera-se que os resultados sejam os mesmos. Caso contrário, não se pode ter confiança no instrumento, porque considera-se não haver consistência nas medidas.

Mais tarde, com o desenvolvimento das teorias de aprendizagem e, conseqüentemente, com as de avaliação, alguns pressupostos assumidos nesta perspectiva, como o da universalidade e o da unidimensionalidade, passam a ser questionados. O primeiro estabelece que resultados iguais em um teste significam desempenhos iguais e que o escore em um teste representa a habilidade que o instrumento se propõe medir. O segundo pressuposto estabelece que um item deve medir um e somente um atributo.

A partir da segunda metade do século passado, foi se tornando cada vez mais claro que se pode concluir muita coisa com base no resultado de apenas um teste, pois sabemos que um único instrumento não dá conta de informar o conjunto das variáveis integrantes do fenômeno a ser estudado, mas apenas alguns de seus aspectos e que os atributos são multidimensionais.

De acordo com Vianna (1997), em 1963, Glaser apresentou uma proposta referindo-se a um instrumento que servisse para a

avaliação formativa e que tivesse como padrão de referência o próprio indivíduo, criando a expressão – testes referenciados a critério. Essa proposta centra-se na instrução individualizada e no uso de objetivos comportamentais (Bloom, 1968).

Na mesma época, movimentos de reforma curricular ocorreram em diversos países europeus e também nos Estados Unidos. Na Holanda, o movimento de reforma do ensino da matemática resultou numa abordagem que passou a ser denominada Educação Matemática Realística (RME).

As ideias do matemático Hans Freudenthal (1905-1990) foram essenciais para o desenvolvimento da RME. Três pilares sustentam essa abordagem: o entendimento do que é matemática, de como a aprendizagem matemática se dá e de como deve ser o ensino de matemática (VAN DEN HEUVEL-PANHUIZEN, 1996).

Para Freudenthal, a matemática é uma atividade humana e não um “estoque de conhecimentos e habilidades nas prateleiras do supermercado” (FREUDENTHAL, 1991, p.15).

Entender a matemática como uma atividade é bem diferente do que entendê-la do modo como está impressa em livros e, de certa forma, em nossas mentes, como um conjunto de produtos finalizados.

De um lado, quando a matemática é tratada como produto, os caminhos traçados pelos matemáticos para seu desenvolvimento e os processos de refinamento pelos quais os conceitos passam no decorrer do tempo são desconsiderados. A crença de que o direito de praticar matemática deve ser reservado apenas para aqueles que produzem “nova matemática” e a compartimentação dos conteúdos matemáticos são alguns fatores que contribuem para esse ponto de vista. Por outro lado, a matemática entendida como atividade humana, assim como a fala, a escrita ou o desenho, pode ser desenvolvida por cada indivíduo. Desse modo, os alunos aprendem matemática fazendo-a, reinventando-a, em uma atividade semelhante à atividade desenvolvida por matemáticos. Essa atividade é chamada de matematização. Nessa perspectiva, o foco

está no processo matemático desenvolvido, no “fazer” matemática e não no resultado obtido, ou seja, no produto final.

Segundo Freudenthal (1971, p. 413), “a matemática como atividade humana é uma atividade de resolução de problemas, de procura por problemas, mas é também uma atividade de organização de um determinado assunto”. Segundo Gravemeijer e Terwel (2000 p. 781, tradução nossa), ao matematizar, busca-se

a *generalidade*: generalizando (olhando para analogia, classificando, estruturando); a *certeza*: refletindo, justificando, provando (usando uma abordagem sistemática, elaborando e testando conjecturas, etc.); a *exatidão*: modelando, simbolizando, definindo (limitando interpretações e validade); e a *concisão*: simbolizando e esquematizando (desenvolvimento de padrão, procedimentos e notações).

Na RME, não apenas o processo matemático é destacado, mas também o meio no qual ele ocorre. De acordo com Freudenthal (1991), “o real” inclui objetos e atividades mentais, e a realidade é uma mistura indissociável de interpretação e de experiência sensual. Os conceitos resultam de processos cognitivos e não o contrário. Por esse motivo, Freudenthal considerava uma inversão antididática iniciar o processo de ensino de matemática com a explicitação de conceitos ou definições, que não fazem sentido algum para o aluno, que não fazem parte das suas realidades matemáticas, que são resultado da atividade matemática de outros (GRAVEMEIJER, 2005).

Buscando evitar a inversão antididática, Freudenthal propõe a reinvenção guiada. Entendendo que não será possível que aluno invente sozinho toda a matemática que levou muito tempo para ser construída, o mesmo autor propõe que o professor, com o auxílio de materiais/ferramentas didáticos, ofereça oportunidades guiadas para que o aluno reinvente a matemática (GRAVEMEIJER, 2005). Os estudantes deixam de ser tomados como meros receptores de matemática já pronta, e passam a ser considerados participantes ativos dentro dos processos de ensino e de aprendizagem nos quais desenvolvem ferramentas matemáticas e *insights*.

A aprendizagem matemática ocorre em diferentes níveis de formalização/complexidade. Com base nisso, é tomada a ideia de modelos emergentes uma vez que, ao lidar com uma determinada situação, os alunos desenvolvem estratégias e ferramentas que podem ser informais e estar conectadas ao contexto do problema. Os alunos constroem “um modelo de” resolução para a situação. Posteriormente, ao lidar com situações semelhantes, o foco pode transferir-se para as relações e estratégias matemáticas. O modelo toma um caráter mais objetivo e começa a tornar-se uma base referencial para um raciocínio matemático mais formal e se torna “um modelo para” resolver/lidar com determinadas situações. Assim, a aprendizagem ocorre à medida que as estratégias utilizadas pelos alunos se tornam mais avançadas/elaboradas e não pela provocação causada pela oferta de problemas cada vez mais complexos (VAN DEN HEUVEL-PANHUIZEN, 2001).

Essas ideias deram origem a seis princípios que visam orientar o trabalho na perspectiva da RME. O resumo das características que descrevem esses princípios foi apresentado por Ferreira (2013, p. 37).

**Quadro 01 - Características dos princípios da RME**

<b>Princípios</b>	<b>Características</b>
(1) Da Atividade	<ul style="list-style-type: none"> <li>- refere-se à interpretação da matemática como atividade humana (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- aprender é uma atividade construtiva (NES, 2009);</li> <li>- as produções dos estudantes são utilizadas para a construção de conceitos (VAN DEN HEUVEL-PANHUIZEN, 2000).</li> </ul>
(2) Da Realidade	<ul style="list-style-type: none"> <li>- a RME tem a função de tornar os alunos capazes de aplicar matemática (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- o processo de matematização ocorre a partir da exploração de contextos ricos (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- fenômenos da realidade devem ser organizados por meio da matemática (NES, 2009);</li> <li>- é importante o uso de contextos reais que sejam significativos e naturais ao aluno como ponto de partida</li> </ul>

	para a sua aprendizagem (WIDJAJA; HECK, 2003).
(3) De Níveis	<ul style="list-style-type: none"> <li>- os alunos passam vários níveis de compreensão (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- começam de seus procedimentos informais e por meio da matematização progressiva e esquematizações avançam para a construção de modelos mais formais (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- os modelos têm de mudar de “modelo de” ao “modelo para” (STREEFLAND, 1991).</li> </ul>
(4) Do Entrelaçamento	<ul style="list-style-type: none"> <li>- domínios matemáticos, como geometria, número, medição e manipulação de dados não são considerados capítulos curriculares isolados, mas fortemente integrados (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- os alunos devem desenvolver uma visão integrada da matemática, bem como flexibilidade para se conectar a diferentes subdomínios e / ou a outras disciplinas (WIDJAJA; HECK, 2003);</li> <li>- a resolução de problemas de contexto ricos significa muitas vezes que se tem de aplicar uma ampla gama de ferramentas matemáticas e entendimentos (VAN DEN HEUVEL-PANHUIZEN, 2000);</li> <li>- a força do princípio entrelaçamento é que traz coerência para o currículo. Este princípio refere-se não só aos diferentes domínios de matemática, mas também podem ser encontradas dentro deles (VAN DEN HEUVEL-PANHUIZEN, 2000).</li> </ul>
(5) Da Interatividade	<ul style="list-style-type: none"> <li>- a aprendizagem matemática não é apenas uma atividade pessoal, mas também uma atividade social (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- os alunos devem ter oportunidades para compartilhar suas estratégias e invenções com outros alunos (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- a interação entre alunos e professores é uma parte essencial na RME porque a discussão e colaboração oportunizam a reflexão sobre o trabalho (WIDJAJA; HECK, 2003).</li> </ul>
(6) De Orientação	<ul style="list-style-type: none"> <li>- os estudantes devem contar com uma oportunidade “guiada” para “reinventar” a matemática (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- o ensino e os programas devem basear-se num conjunto coerente de trajetória de ensino-aprendizagem a longo prazo (VAN DEN HEUVEL-PANHUIZEN, 2010);</li> <li>- os alunos precisam de espaço para construir conhecimentos matemáticos e ferramentas por si só. Para alcançar isso, os professores têm de proporcionar aos alunos um ambiente de aprendizagem em que este processo de construção possa surgir (VAN DEN HEUVEL-</li> </ul>

PANHUIZEN, 2000).
-------------------

**Fonte:** Ferreira (2013 p.37)

A avaliação atende aos princípios para essa abordagem. Assim, a avaliação se sustenta nos mesmos três pilares que sustentam a RME. Van den Heuvel-Panhuizen (1996) destaca duas características principais da avaliação na RME: sua natureza didática e a importância do papel desempenhado pelos problemas escolhidos.

Para essa autora, a natureza didática da avaliação pode ser observada não apenas no seu propósito, mas também em outros aspectos, como o conteúdo, os procedimentos e as ferramentas. O principal propósito da avaliação na RME é obter informações que contribuam para os processos de ensino e de aprendizagem e que auxiliem os envolvidos no processo de tomada de decisões educacionais. Para isso, não apenas o aluno e seus processos de aprendizagem são avaliados, mas também a instrução. A avaliação deve ser parte dos processos de ensino e de aprendizagem, e não uma interrupção do mesmo. Quanto mais próxima for a avaliação da educação que acontece em sala de aula, maiores serão as chances de que ela contribua para o desenvolvimento educacional.

O conteúdo das provas deve abranger todos os componentes curriculares, as relações entre eles e todos os níveis de compreensão. O foco principal não está nos resultados, mas sim nos processos que levam aos resultados. A avaliação deve fornecer *insights* das atividades de matematização dos alunos. A avaliação deve fazer parte de cada fase dos processos de ensino e de aprendizagem. Desse modo, não cabe apenas olhar para o que o aluno já aprendeu, mas também para o que ele ainda pode aprender. Em relação a esse aspecto, o professor exerce um papel fundamental, já que é ele quem decide o que (quais conteúdos), como (quais instrumentos, tarefas) e quando (em que momento) será avaliado em relação a esse aspecto (VAN DEN HEUVEL-PANHUIZEN, 1996).

As ferramentas devem ser diversificadas de modo a fornecer o máximo de informações possível a respeito dos alunos. Frequentemente elas devem ser indistinguíveis das utilizadas para iniciar os processos de aprendizagem, devem expor o processo de aprendizagem e fornecer *insight* dentro do repertório de conhecimentos dos alunos, habilidades e *insights* em dado momento. Uma distância é claramente mantida em relação a uma abordagem psicométrica em que uma superperseguição de objetivo e avaliação confiável frequentemente ocorre a custo do conteúdo.

Os aspectos da natureza didática da avaliação na RME não podem ser considerados independentemente um do outro, pois há uma ligação recíproca entre eles (VAN DEN HEUVEL-PANHUIZEN, 1996). Esses aspectos e a integração entre eles também podem ser percebidos em alguns dos nove princípios para avaliação em sala de aula propostos por De Lange (1999, p. 10):

1. O principal objetivo da avaliação em sala de aula é promover a aprendizagem.
2. A matemática escolar deve ser integrada a situações que fazem parte do mundo real do aluno.
3. Métodos de avaliação devem ser tais que permitam que estudantes revelem o que sabem mais do que aquilo que eles não sabem.
4. Um plano de avaliação equilibrada deve incluir múltiplas e variadas oportunidades (formatos) para os alunos mostrarem e documentarem em suas realizações.
5. As tarefas devem operacionalizar todas as metas do currículo, e não apenas as mais facilmente avaliadas. Os padrões de desempenho são úteis para este fim, incluindo a indicação dos diferentes níveis do pensamento matemático.
6. Os critérios de classificação devem ser públicos e consistentemente aplicados; e devem incluir exemplos de classificações anteriores mostrando um trabalho exemplar e de um trabalho que não seja tão exemplar.
7. O processo de avaliação, incluindo a pontuação e classificação deve ser aberto aos estudantes.
8. Os estudantes devem ter oportunidades de receber *feedback* real sobre o seu trabalho.
9. A qualidade de uma tarefa não é definida pela sua acessibilidade a pontuação objetiva, confiabilidade ou validade, no sentido tradicional, mas pela sua autenticidade, equidade, na medida que atende aos princípios acima mencionados.

Em relação à outra característica mencionada: o papel essencial que os problemas exercem, Van den Heuvel-Panhuizen (1996) destaca que as tarefas da avaliação não são diferentes das propostas em sala de aula. Não ter tarefas específicas para a avaliação é uma característica marcante na RME.

Começando com um problema particular, os alunos desenvolvem ferramentas matemáticas e conhecimentos com os quais eles podem resolver novos problemas. Estes problemas são vistos como situações que requerem uma solução, o que pode ser alcançado por meio de esquematização, organização e o processamento dos dados, ou, por outras palavras, pela matematização. Um aspecto essencial é a reflexão sobre as atividades matemáticas que podem permitir um aumento no nível (VAN DEN HEUVEL-PANHUIZEN, 1996, p.88)

Ainda de acordo com Van den Heuvel-Panhuizen (1996), os problemas podem também divergir consideravelmente em relação à complexidade e ao nível e podem envolver tanto situações da vida real, quanto situações puramente matemáticas. A autora descreve as características de bons problemas para a avaliação.

**Quadro 02 - Características de bons problemas para a avaliação**

<b>ACESSÍVEL</b>	<ul style="list-style-type: none"> <li>- Ser tão claro quanto possível, os alunos precisam compreendê-lo, refletir a respeito do assunto nele envolvido, se colocarem na situação. Isso não significa que eles devem conhecer de antemão alguma solução para ele.</li> <li>- Permitir que os alunos possam abordá-lo de alguma maneira.</li> </ul>
<b>SIGNIFICATIVO</b>	<ul style="list-style-type: none"> <li>- Ser acessível.</li> <li>- Convidativo, interessante, estimulante e desafiador.</li> <li>- Valer a pena. Os alunos precisam ter um motivo para resolvê-lo. O problema deve exigir uma solução.</li> <li>- Permitir que os alunos moldem a situação, se posicionem, tomem decisões e apresentem soluções autênticas.</li> <li>- Permitir que os alunos apresentem ou pensem em questões próprias.</li> <li>- Do ponto de vista da avaliação, eles devem</li> </ul>

	refletir objetivos importantes.
<b>ELÁSTICO/FLEXIVEL</b>	- Permitir que seja resolvido de diversas maneiras e em diferentes níveis, o que significa que deve haver espaço para as construções dos próprios alunos.
<b>TRANSPARENTE</b>	- Permitir ao estudante mostrar o nível em que se encontra. A fim disso, deve: i) exigir mais do que lembrar um fato ou a reprodução de uma habilidade, ii) ter uma componente educativa, o que significa que os alunos e professores vão aprender com a tentativa de respondê-lo e iii) ser abertos em certa medida, o que significa que várias respostas podem ser possíveis. (Lembrando que isso deve servir para que o professor possa planejar suas ações de modo a auxiliar o estudante para que ele possa, em outro momento, atingir outros níveis.).
<b>INFORMATIVO</b>	- Fornecer ao professor um máximo de informações sobre o conhecimento dos alunos, visão e habilidades, incluindo suas estratégias. - Permitir que haja espaço para as construções dos próprios alunos, o que significa que o problema deve ser de um tipo que pode ser resolvido de diversas maneiras e em diferentes níveis.

**Fonte:** A autora

Outros aspectos em relação às tarefas de avaliação são discutidos por De Lange (1999). O autor considera três níveis relativos à complexidade cognitiva das tarefas, entendendo que, ao fazer matemática, o aluno recorre a diversas competências ao mesmo tempo. Os três níveis são:

**Nível I – Reprodução, definições, cálculos.** Para resolver tarefas do nível I, o aluno recorre a competências como representar, reconhecer equivalências, recordar objetos matemáticos e propriedades, realizar procedimentos de rotina, aplicar algoritmos padrão, desenvolver habilidades técnicas e operacionalizar declarações e

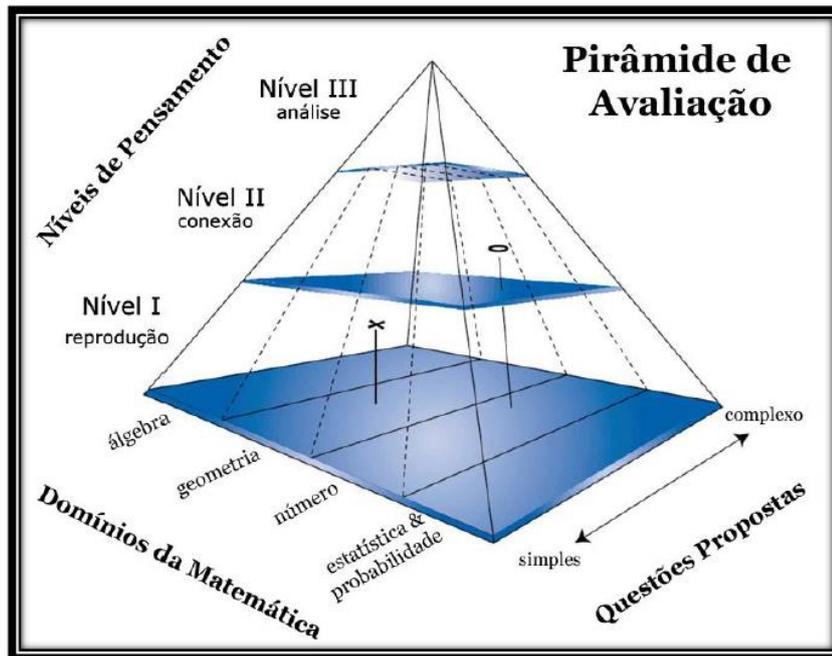
expressões que contêm símbolos e fórmulas em formato padrão (DE LANGE, 1999).

**Nível II – Conexões e integração para resolução de problemas.** As tarefas do nível II demandam competências como fazer conexões entre as diferentes vertentes e domínios da matemática, integrar informações para resolver problemas nos quais os alunos podem escolher a estratégia e as ferramentas matemáticas que utilizarão, lidar com diferentes formas de representação, distinguir e relacionar diferentes declarações, tais como definições, reclamações, exemplos, afirmações condicionadas, e de prova, decodificar e interpretar linguagem simbólica e formal e entender suas relações com a linguagem natural (DE LANGE, 1999).

**Nível III – Matematização, pensamento matemático, generalização e perspicácia.** Envolve competências tais como analisar, interpretar, desenvolver modelos e estratégias, criticar o modelo, apresentar argumentos matemáticos incluindo provas e generalizações, refletir sobre o processo, resolver e também propor problemas. Os alunos devem ainda se comunicar de diferentes formas (por exemplo, oralmente, por escrito, usando visualizações). Diz respeito a matematizar situações (DE LANGE, 1999).

Os três níveis são representados em sua pirâmide de avaliação.

**Figura 1** – Pirâmide de Avaliação proposta por De Lange (1999)



**Fonte:** Ferreira (2013, p. 61)

As demais dimensões da pirâmide dizem respeito ao conteúdo matemático ou ao domínio de grandes ideias e ao grau de dificuldade dos itens. A pirâmide serve para dar uma imagem visual equilibrada da quantidade relativa de itens necessários para representar a compreensão de um estudante de matemática.

Em resumo, isso significa que são apresentadas aos alunos situações-problema que eles podem imaginar, interpretar vestígios 'realísticos', sendo essas características voltadas à expressão holandesa 'zich REALISERen', significando 'imaginar'. É essa ênfase em fazer algo real em sua mente que deu nome à RME. Portanto, na RME, problemas apresentados aos alunos podem vir do mundo real, mas também do mundo de fantasia, dos contos de fadas, ou do mundo formal da matemática, uma vez que os problemas são experimentalmente reais na mente do aluno.

## DESENLACE

Conforme a teoria de avaliação evoluiu e passou a se estruturar de acordo com as teorias de aprendizagem, a ideia de que os indivíduos podem desenvolver a capacidade de aprender começa a ser considerada, em divergência com o entendimento da inteligência como uma característica inata e imutável no ser humano, que determinava a capacidade de aprender de cada um. O propósito da avaliação passa a ser fornecer elementos para orientar a aprendizagem. E a integração entre ensino, aprendizagem e avaliação é levada em conta. Com isso, a validade da avaliação pode ser uma forma de proporcionar uma reflexão equilibrada do que os alunos podem ser capazes de fazer (LESH e LEMON, 1992). Considerando a concepção de avaliação didática adotada na abordagem RME, foi elaborado um quadro com o qual se pretende apresentar características dela e da concepção clássica<sup>9</sup> de avaliação.

**Quadro 03** – Algumas características do caráter de validade da concepção clássica de avaliação e da concepção de Avaliação Didática (RME)

Na concepção clássica de avaliação	Na concepção da Avaliação Didática
<ul style="list-style-type: none"> <li>- A validade não é um atributo geral e pode existir em diferentes graus em relação a determinado propósito e para um grupo particular, com características específicas.</li> <li>- Uma avaliação é válida na medida em que mede o que ela se propôs a medir. A validade refere-se ao que o teste mede.</li> <li>- De acordo com Vianna (1997), a avaliação inicialmente teve seus fundamentos na psicometria. Nesse contexto, buscava-se construir instrumentos com o intuito de medir</li> </ul>	<ul style="list-style-type: none"> <li>- Um dos aspectos da avaliação didática é que estudantes e professores participam na produção e uso das informações, por conseguinte, uma tarefa só pode ser válida para um grupo particular, com características específicas, e mais, em um tempo específico.</li> <li>- Na perspectiva da RME, os instrumentos de avaliação não têm a função de fornecer medidas.</li> <li>- Para discutir o significado de validade no contexto da avaliação didática, é relevante colocar, também, em pauta que a avaliação</li> </ul>

<sup>9</sup> Clássico aqui tomado como a forma que se apresenta habitual e consagrada na educação escolar.

a inteligência, entendendo-a como uma característica inata e imutável no ser humano, determinando a capacidade de aprender de cada um.

- A avaliação tradicional se apoiou em ideias que tiveram seu momento na década de 20. Como, por exemplo, a concepção de que a aprendizagem seria linear e sequencial e de que, para ensinar os conhecimentos mais complexos, seria necessário atomizá-los. Encontramos, ainda hoje, muitos professores que agem em acordo com essas ideias, ou por acreditarem nelas ou por não pensarem no assunto, apenas seguem a tradição.

- A construção dos testes assume um modelo comportamental de aprendizagem. Os resultados são interpretados como base em normas, isso significa que os desempenhos dos indivíduos são comparados com os desempenhos dos demais indivíduos do grupo, de modo a classificá-los. Assim, quanto maior for a amplitude de variação dos resultados encontrados melhor a distinção entre eles. Por isso, itens fáceis ou difíceis são eliminados após um pré-teste.

- Tradicionalmente, segundo Vianna (1997), a fidedignidade era considerada como uma condição necessária para a validade em um teste e dizia respeito à reprodutibilidade - se um teste é aplicado ao mesmo grupo um grande número de vezes, espera-se que os resultados sejam os mesmos, desde que o grupo não se modifique. Se isso não ocorresse, as

didática precisa ser planejada de modo a contemplar os três pilares da RME: (1) olhar as várias ideias, os significados e os pontos de vista sobre a matéria; (2) a forma como essa matéria deve ser ensinada e (3) o modo pelo qual o processo de aprendizagem se desenvolve. Juntos, eles determinam o quê, o porquê e como influenciam a avaliação. Com isso, quando se elabora, se executa e se reflete nos seus resultados, tem-se uma boa aproximação da função formativa da avaliação didática. Assim, a aprendizagem se destaca como é entendida na RME.

- A RME não é tomada como uma teoria fixa e acabada, mas como uma abordagem de ensino de um domínio específico (a Matemática), um projeto em andamento, cujo foco é a aprendizagem, e que, de acordo com Van den Heuvel-Panhuizen e Drijvers (2014), fundamenta a perspectiva para a avaliação denominada avaliação didática. Nela, ao contrário daquela fundamentada na psicometria, não se buscam instrumentos para medir a inteligência, e sim instrumentos que possam oportunizar a aprendizagem.

- Na RME, a natureza descontínua dos processos de aprendizagem, os quais envolvem atividades dos próprios alunos, seus níveis de compreensão e matematização, não permite a concepção de que a aprendizagem é linear, sequencial e que, para ensinar os conhecimentos mais complexos, seria necessário atomizá-los. Com isso, um critério essencial para uma boa tarefa de avaliação é que ela possa contribuir

medidas derivadas do teste eram consideradas inconsistentes.

- Em sua forma clássica, a validade de conteúdo de teste diz respeito à representatividade dos conhecimentos e comportamentos adquiridos durante o processo educacional (VIANNA, 1978). Em acordo, Messick (1979) afirma que existem duas facetas principais para a validade de conteúdo: a relevância e a cobertura do conteúdo. O autor afirma ainda que a validade de conteúdo é uma das bases principais para julgar a relevância do teste para o domínio comportamental no que diz respeito a quais inferências devem ser desenhadas.

- Outros dois tipos de validade são frequentemente encontrados na literatura, a validade concorrente e a validade preditiva. De acordo com a teoria clássica de avaliação, a validade concorrente é a relação entre os escores de um teste e os escores de um critério que medem o mesmo desempenho. Os escores devem ser obtidos simultaneamente. Essa é a principal diferença entre a validade concorrente e a preditiva. Na validade preditiva, os escores dos testes são obtidos independentemente das outras medidas do desempenho (critério).

para o envolvimento ativo dos alunos, de modo que eles se sintam "donos" da tarefa, assumindo uma **responsabilidade** maior com sua resolução.

O que é característico de todos os cuidados que podem ser tomados é que, por um lado, eles muitas vezes fazem as tarefas mais acessíveis para os alunos; por outro, é que as páginas da prova podem revelar ao professor muito sobre o nível de compreensão de seus alunos e das estratégias aplicadas.

- Na avaliação didática, os resultados não são os obtidos na resolução das tarefas de avaliação, mas toda a sequência da própria resolução, as estratégias escolhidas, os procedimentos desenvolvidos, enfim, as maneiras de os alunos lidarem com as tarefas, sem a pontuação alta nas respostas dadas. Pontuar o correto ou o incorreto é insuficiente. Isso acarreta uma mudança radical relativa a uma prática em que a classificação tem sido a ação prioritária, e, com isso, são descartadas todas as informações que não são consideradas "pontuáveis".

Em uma avaliação didática, a riqueza das resoluções e respostas dos alunos e a possibilidade de distintos aspectos presentes nelas não permitem a utilização do Princípio do Terceiro Excluído – está correto ou incorreto - sem outra alternativa. Isso implica analisar, em cada resolução, o que os alunos poderiam ter entendido para apresentá-la.

- A máxima da avaliação baseada na psicometria - "corrigir como uma máquina faria" - é completamente

sem sentido na avaliação didática, pois o princípio de que ela deve contribuir para a aprendizagem é incompatível com o desejo de que os alunos encontrem sempre os mesmos resultados em testes repetidos. O que se adota é um critério de equidade, que implica em considerar as diferenças entre os alunos e oportunizar que todos mostrem o que sabem, até porque o fato de alunos tirarem a mesma "nota" não garante que tenham o mesmo conhecimento do assunto contido na tarefa.

De acordo com Freudenthal, é a maneira como a resposta a uma pergunta é encontrada que determina a realização de um dado nível, e não simplesmente a capacidade de responder à pergunta.

- Um pesquisador ou um professor, pode encaminhar a resolução de uma tarefa de modo bem diferente de um aluno. Isso implica aceitar que o nível da resposta não é determinado pela tarefa, mas principalmente pelo revolvedor.

- Outros tipos de validade estão presentes na literatura da área: de conteúdo, de construto, preditiva, concorrente. De certa forma, todas situadas na mensuração psicológica. Na RME, a foco é na dimensão pedagógica, e a busca da certeza psicométrica é abandonada dando lugar a uma abordagem hermenêutica. Nela, a intenção é elaborar uma interpretação coerente do desempenho, que pode ser constantemente revista, criando oportunidade para uma prática mais humana, na mesma categoria que Freudenthal coloca a matemática –

uma atividade humana.
-----------------------

**Fonte:** a autora

As definições apresentadas nos dicionários com relação à origem da palavra e de seu significado filosófico convergem, mas são definições muito gerais que apontam a utilidade ou serventia de um instrumento ou de qualquer outra coisa em relação a sua finalidade. Essas definições dizem muito pouco a respeito da validade na avaliação educacional, no entanto apontam que, para identificar aspectos do caráter de validade em uma determinada perspectiva de avaliação, é necessário o esclarecimento da finalidade dessa avaliação e de outros aspectos relacionados a ela.

Portanto, para configurar o caráter de validade na RME, faz-se necessário discutir qual é a finalidade da avaliação na RME. Além da finalidade, outros aspectos que se relacionam com a finalidade serão retomados.

A avaliação na RME não se apoia em fundamentos psicométricos e sim nos mesmos três pilares que sustentam a abordagem, ou seja, os pontos de vista sobre a matemática, a forma como ela deve ser ensinada e a maneira como se aprende. Segundo Van den Heuvel-Panhuizen (1996), os pontos de vista a respeito dos processos de aprendizagem são os que particularmente influenciam a avaliação na RME.

Na RME, a matemática é vista como uma atividade humana. Os processos de aprendizagem levam em consideração as atividades dos próprios alunos, seus níveis de compreensão e matematização, bem como a natureza descontínua dos processos de aprendizagem. O ensino é pensado a partir dos contextos, modelos, contribuições próprias dos alunos, reflexão, interação e integração das vertentes de aprendizagem.

Conforme discutido anteriormente, o principal propósito ou finalidade da avaliação na RME é obter informações que contribuam para os processos de ensino e de aprendizagem e que auxiliem os envolvidos no processo de tomada de decisões educacionais. Portanto, para que as tarefas tenham validade na RME, elas devem atender a essa finalidade. O

caráter **informativo** de bons problemas na RME indica isso, já que sua principal característica é fornecer uma “imagem”, a mais completa possível a respeito do aluno. É importante também lembrar que a imagem que se busca do aluno não é apenas a acerca do que ele já aprendeu, ou seja, os produtos dos processos de aprendizagem, uma vez que a aprendizagem não para em situações de avaliação. Logo, as informações devem se referir aos processos de aprendizagem dos estudantes durante a avaliação. Nesse sentido, é importante que os problemas oportunizem a aprendizagem e que suscitem processos de pensamento; que os estudantes sejam convidados a refletir e revelar aspectos dos processos empregados nas soluções. Para isso, devem se envolver ativamente com a avaliação, conseqüentemente os problemas precisam ser **significativos**. Devem permitir construções próprias dos estudantes. Apenas reproduzir procedimentos não basta.

É necessário entender que problemas informativos podem ser resolvidos de diversas maneiras, que podem ter mais de uma resposta correta, ou que, em alguns casos, a resposta correta pode não ser determinada. Problemas que permitem uma só resposta correta levarão quase sempre a uma situação na qual o professor apenas poderá verificar se o estudante sabe o procedimento para resolver ou não, ou seja, se errou ou acertou a questão. E se o estudante não conhecer o procedimento, não poderá mostrar outras coisas que conhece. Problemas desse tipo devem ser evitados na medida do possível. Os problemas devem ser **elásticos/flexíveis**. Problemas elásticos, embora sejam um tabu na avaliação psicométrica, na RME, são bem vistos pela riqueza de informações que podem fornecer. Por um lado, eles podem se tornar mais **acessíveis** aos alunos, já que permitem soluções em diferentes níveis; por outro, podem revelar ao professor o nível de compreensão em que o estudante se encontra, ou seja, o aspecto de **transparência** das tarefas.

Além disso, os problemas selecionados devem fornecer evidências para que se possa fazer inferências adequadas a respeito da aprendizagem dos estudantes. Não será possível fazer inferências acerca

de um determinado nível de pensamento do aluno, se o professor não oferecer ao estudante tarefas que possibilitem isso (quinto princípio de De Lange). As tarefas devem permitir que os diferentes níveis de pensamento dos alunos sejam demonstrados. Por isso, não é a facilidade ou não da tarefa que deve ser levada em conta no momento de selecioná-la. Ao fazer matemática, os alunos recorrem simultaneamente a diversas habilidades que De Lange buscou agrupar em três níveis. Se a demanda cognitiva da tarefa for apenas reprodução, memorização, então o professor só poderá fazer inferências em relação a essas habilidades desenvolvidas por seus alunos, e nada mais. No entanto, é importante lembrar que os níveis não devem ser considerados como absolutos. Um determinado problema pode não "incluir" o nível de pensamento cognitivo pretendido, já que é a maneira que a resposta a uma pergunta é encontrada que determina a realização de um dado nível. Assim, é o estudante que determina o nível da tarefa e não a tarefa *per se*. Apesar disso, os alunos devem ter a oportunidade de responder em um nível particular, e isso requer problemas apropriados. Portanto, os níveis devem funcionar como um modelo global para dar suporte para outras instruções a partir das informações obtidas.

No entanto, não basta fornecer problemas adequados, é necessário que o professor olhe para as soluções com outro olhar; não buscando distinguir entre o certo ou errado, mas analisando os processos de pensamento dos seus alunos. Problemas que atendem às características discutidas muitas vezes não permitirão tal distinção, mas pedirão mais informações a partir das respostas apresentadas. Apenas classificar os problemas é não fazer justiça à riqueza de informações que eles podem fornecer. Cabe ao professor tentar se colocar no lugar do aluno e se perguntar qual pode ter sido o seu raciocínio.

A "objetividade" na correção não faz sentido nessa abordagem. A qualidade da avaliação, conforme o nono princípio de De Lange, deve ser definida pela equidade e não pela objetividade e normalização, ao contrário do que é defendido na psicometria. Isso

significa considerar as diferenças entre os alunos e oferecer-lhes oportunidades de mostrar o que podem fazer. Em contraste com a indicação psicométrica, em que os professores deveriam desconsiderar qualquer conhecimento que tinham dos alunos, ao corrigir os seus trabalhos, agora é cada vez mais necessário lançar mão desse conhecimento. Além de permitir aos alunos mais espaço para interpretar e resolver os problemas à sua maneira, pode ser considerado necessário, pelo bem da equidade, fornecer aos alunos mais assistência. Isso é algo que, até recentemente, era inconcebível.

Na RME, a busca da certeza psicométrica é abandonada, dando lugar a uma abordagem hermenêutica. Nela, é dada importância ao papel atribuído ao julgamento humano, na intenção de desenvolver interpretações integradas e coerentes com base em todas as evidências relevantes. Interpretações feitas anteriormente podem ser constantemente revistas, já que a inconsistência no desempenho dos alunos em tarefas não invalida a avaliação. Em vez disso, ela se torna um enigma empírico a ser resolvido por meio da procura de uma interpretação mais abrangente ou elaborada que explicaria a incoerência ou articularia a necessidade de provas adicionais. A ideia de que a avaliação deve contribuir para a aprendizagem não é compatível com a exigência de que os alunos precisam encontrar os mesmos resultados em testes repetidos.

Assim, o caráter de validade na avaliação didática depende das ações desenvolvidas tanto pelo professor quanto pelos alunos nos processos de ensinar e aprender matemática na escola. Então, tomando a expressão validade como indicativo da utilidade ou da característica de que algo se encontra em condições de produzir os efeitos dele esperados e considerando que a avaliação deve ser uma oportunidade de aprendizagem, então a validade de uma avaliação consiste em servir a esse propósito, ou seja, oportunizar a aprendizagem. É a esse significado de validade que este trabalho diz respeito.

## REFERÊNCIAS

ABBAGNANO, Nicola. **Dicionário de filosofia**. São Paulo: Mestre Jou, 1982.

BLACK, Paul. **Os professores podem usar a avaliação para melhorar o ensino?** Tradução de Fernando Zan Vieira. *Práxis Educativa*, Ponta Grossa, v.4, n.2, jul.-dez. 2009. Disponível em: <<http://www.periodicos.uepg.br>>. Acesso em: 15 jun.2017. Tradução de: Can teachers use assessment to improve learning?

CUNHA, Antônio Geraldo da, **Dicionário etimológico nova fronteira da língua portuguesa**. Rio de Janeiro: Editora Nova Fronteira, 1982.

DE LANGE, J. **Framework for classroom assessment in mathematics**. Madison: WCER, 1999.

FERREIRA, Aurélio Buarque de Holanda. **Novo dicionário Aurélio da língua portuguesa**. Curitiba: Ed. Positivo, 2009.

FERREIRA, Pamela Emanuelli Alves. **Enunciados de Tarefas de Matemática: um estudo sob a perspectiva da Educação Matemática Realística**. 2013. 121f. Tese (Programa de Pós-Graduação em Ensino de Ciências e Educação Matemática) – Universidade Estadual de Londrina, Londrina, 2013.

FREUDENTHAL, H. Geometry between the devil and the deep sea. **Educational Studies in Mathematics**, v. 3, n. 3-4, p. 413-435, 1971.

\_\_\_\_\_. **Revisiting Mathematics Education**. Netherlands: Kluwer Academic Publishers, 1991.

GRAVEMEIJER, K. P. E.; TERWEL J. Hans Freudenthal: a mathematician on didactics and curriculum theory. **Journal of Curriculum Studies**, v. 32, n. 6, p. 777-796, nov-dez. 2000.

GRAVEMEIJER, K. P. E. O que torna a Matemática tão difícil e o que podemos fazer para o alterar?. **Educação matemática: caminhos e encruzilhadas**. Lisboa: APM, p. 83-101. 2005.

LESH, Richard A.; LAMON, Susan J. (Ed.). Trends, Goals, and Priorities in Mathematics Assessment. In: \_\_\_\_\_. **Assessment of Authentic Performance in School Mathematics**. Washington: Library of Congress, 1992. p.5.

HOUAISS, A. **Dicionário Eletrônico da Língua Portuguesa**. Rio de Janeiro: Objetiva, 2009. CD-ROM.

MESSICK, Samuel. **Test Validity and the Ethics of Assessment**. 1979. Disponível em: <<http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.1979.tb01178.x/epdf>>. Acesso em: 26 abr. 2017.

MORROW JUNIOR, J. R. et al. **Medida e avaliação do desempenho humano**. 4. ed. Porto Alegre: Artmed, 2013. p.121-135.

VAN DEN HEUVEL-PANHUIZEN, M. V. D. **Assessment and Realistic Mathematics Education**. Utrecht: CD-β Press/Freudenthal Institute, Utrecht University. 1996.

\_\_\_\_\_. Learning-teaching trajectories with Intermediate attainment targets. In: VAN DEN HEUVEL-PANHUIZEN, Marja (Ed.). **Children learn mathematics: a learning-teaching trajectory with intermediate attainment targets for calculation with whole numbers in primary school**. Groningen, The Netherlands: Wolters Noordhoff, 2001.

\_\_\_\_\_. Realistic Mathematics Education in the Netherlands. In: ANGHILERI, Julia (Ed.), **Principles and practice in arithmetic teaching** Buckingham/Philadelphia: Open University Press, p. 49-63, 2001.

VAN DEN HEUVEL-PANHUIZEN, M., & DRIJVERS, P. Realistic Mathematics Education. In S. Lerman (Ed.), **Encyclopedia of Mathematics Education** (pp. 521-525). Dordrecht, Heidelberg, New York, London: Springer, 2014.

VIANNA, Heraldo Marelím. Validade dos testes. In: \_\_\_\_\_. **Testes em educação**. 3. ed. São Paulo: Ibrasa, 1978. p.171-185.

VIANNA, Heraldo M. Avaliação educacional e seus instrumentos: novos paradigmas. In: SOUSA, Eda C.B. Machado de (Org.). **Técnicas e Instrumentos de Avaliação**. Brasília: Universidade de Brasília, 1997. p. 36-84.

WILIAM, Dylan. **The validity of teachers' assessments**. 1998. Disponível em: <<http://www.leeds.ac.uk/educol/documents/000000792.htm>>. Acesso em: 15 mai. 2017.

## **APÊNDICES**

**APÊNDICE A** – Resultado da busca da palavra validade em dicionários

<p>Dicionário de Língua Portuguesa Houaiss</p> <p>HOUAISS, A. <b>Dicionário Eletrônico da Língua Portuguesa.</b> Rio de Janeiro: Objetiva, 2009. CD-ROM.</p>	<p>VALIDADE</p>	<p>Qualidade ou condição de algo que se encontra em condições de produzir os efeitos dele esperados; valimento.</p> <p>Etimologia – válido + dade</p>
	<p>VÁLIDO</p>	<p>Que tem valor; valioso. Que tem eficácia, é capaz de produzir efeitos legais. Que foi produzido respeitando as formalidades legais. Que tem embasamento bem fundado; correto, certo. Que é logicamente correto, seja do ponto de vista formal, seja do ponto de vista interpretativo. Que é apropriado ao fim que se destina.</p> <p>Etimologia – lat. validus, a, um 'forte, robusto, vigoroso; sólido; saudável'.</p>
	<p>VALOR</p>	<p>Determinação quantitativa obtida por cálculo ou mensuração; número. Utilidade, préstimo, serventia. Legitimidade, valia. Importância, destaque em uma escala comparativa. Importância estabelecida por convenção. Reconhecimento, importância, consideração. Série de traços culturais, ideológicos, institucionais, morais etc. definidos de maneira sistemática ou em sua coerência interna. Conjunto de princípios ou normas que, por corporificar um ideal de perfeição ou plenitude moral, deve ser buscado pelos seres humanos.</p>
<p>Dicionário de Língua Portuguesa Aurélio</p> <p>FERREIRA, Aurélio Buarque de Holanda. <b>Novo dicionário Aurélio da língua portuguesa.</b> Curitiba:</p>	<p>VALIDADE</p>	<p>Qualidade ou condição de válido; validabilidade. 2. Legitimidade, valência, valimento, valor. 3. Lóg. Qualidade do que é válido em todas as acepções.</p> <p>De válido + dade.</p>

Ed. Positivo, 2009.		
Dicionário de Filosofia Nicola ABBAGNANO	VALIDADE	Utilidade ou eficiência de um meio ou de um instrumento qualquer. Nesse sentido, Dewey afirmou que as proposições, como meios <i>processivos</i> para conduzir uma pesquisa, não são verdadeiras nem falsas, mas apenas válidas (sólidas, eficientes) ou inválidas (débeis, inadequadas) ( <i>Logic</i> , XV; trad. it., pp. 382-83). É a esse significado de V. que se apela sempre que se usa a expressão <i>válido para</i> . O que se segue ao <i>para</i> é o fim ou a função em relação à qual se considera eficiente o instrumento, o meio ou a condição de que se trata. P. ex., um bilhete de viagem é <i>válido para</i> determinado percurso; determinada organização é <i>válida para</i> determinadas funções, etc.
ABBAGNANO, Nicola. <b>Dicionário de filosofia</b> . São Paulo: Mestre Jou, 1982.	VALOR	Em geral, o que deve ser objeto de preferência ou de escolha. Desde a Antiguidade essa palavra foi usada para indicar a utilidade ou o preço dos bens materiais e a dignidade ou o mérito das pessoas. Contudo, esse uso não tem significado filosófico porque não deu origem a problemas filosóficos. O uso filosófico do termo só começa quando seu significado é generalizado para indicar <i>qualquer objeto</i> de preferência ou de escolha, o que acontece pela primeira vez com os estóicos, que introduziram o termo no domínio da ética e chamaram de V. os objetos de escolha moral. Isso porque eles entendiam o bem em sentido subjetivo (v. BEM, 2). podendo assim considerar os bens e suas relações hierárquicas como objetos de preferência ou de escolha. Por V., em geral, entenderam "qualquer contribuição para uma vida segundo a razão" (DIÓG. L-, Vil, 105), ou,

		<p>como diz Cícero, "o que está em conformidade com a natureza ou é digno de escolha (<i>selectionedignam</i>): (<i>De finibus</i>, III, 6, 20). Por "estar em conformidade com a natureza", entendiam o que deve ser escolhido em todos os casos, ou seja, a virtude; como "digno de escolha", entendiam os bens a que se deve dar preferência, como talento, arte, progresso, entre as coisas do espírito; saúde, força, beleza entre as do corpo; riqueza, fama, nobreza, entre as coisas externas (DIÓG. L, VII, 105-06). A divisão entre V. obrigatórios e V. preferenciais será mais tarde expressa como divisão entre V. intrínsecos ou finais e valores extrínsecos ou instrumentais.</p> <p>Etimologia – lat. tar. valore 'id.'</p>
<p>Dicionário Etimológico Nova Fronteira da Língua Portuguesa</p> <p>CUNHA, Antônio da. Geraldo da.</p> <p><b>Dicionário etimológico nova fronteira da Língua Portuguesa.</b> Rio de Janeiro: Editora Nova Fronteira, 1982.</p>	VALIDADE	validAR vb. 'dar validade a , tornar válido' 1813. Do fr. valider, deriv. Do lat. tard. Validare.
	VÁLIDO	adj. 'orig. sadio, forte' 'ext. que tem valor ou valia' XVI. Do lat. Validus.
	VALOR	sm. 'audácia, vigor' 'mérito, importância' 'preço' XIII.

**APÊNDICE B** – Inventário de fragmentos com alguma referência à validade encontrados nos autores estudados

<b>Vianna (1978; 1997a; 1997b; 2002)</b>
A validade e a fidedignidade são consideradas requisitos básicos de um teste. Somente quando essas características estão presentes é possível considerar o teste um instrumento hábil para fins de verificação e avaliação. A validade de conteúdo, estabelecida a partir de um universo de conhecimentos e comportamentos, que se supõe o examinando deva ter adquirido após a sua experiência educacional, é o mais importante dos tipos de validade a "introduzir-se" num teste de desempenho escolar.
A afirmação de que um teste é válido quando mede aquilo que pretende medir é por demais genérica e exige maiores explicitações. A validade não é uma característica geral, não é um problema de "ser" ou "não ser", não é uma questão de "tudo" ou "nada", pois a validade existe em diferentes graus. A validade refere-se <i>ao que</i> um teste mede; conseqüentemente, a palavra validade não pode ser usada isoladamente e apresentada abstratamente. Um teste não é válido de modo geral, mas sim em relação a determinado propósito e para um grupo particular com características específicas.
O problema da determinação do grau de validade de um teste é bem mais complexo que o da fidedignidade, pois depende, em grande parte, da finalidade do teste, da interpretação que se dará aos escores e do seu uso.
A validade é uma característica relativa e existe em diferentes graus. Sua problemática não está condicionada em termos de tudo ou nada. Um teste pode ser válido para certos fins e determinado grupo, mas não o ser para outros fins e indivíduos. A validade não é uma função do teste apenas, mas do uso que dele se faz.
A validade é uma característica complexa e não existe isoladamente. Diferentes tipos de validade coexistem num teste e se interligam para formar um todo, onde este ou aquele tipo de validade pode predominar.
A validade não é um elemento absoluto; depende de uma pluralidade de fatores; além do mais, não é uma característica geral, pois um teste só é válido para determinado fim.
A seleção de um critério fidedigno e válido é fundamental no processo de validação, pois é o critério que define o que o teste mede.
A interpretação da validade exige o cálculo de diferentes coeficientes; entretanto, para fins de aconselhamento e orientação, o uso de tabelas de expectativa é o mais recomendável.
A fidedignidade não é o único elemento a considerar na validade de um teste; é condição necessária, mas não suficiente para que um teste seja válido.

<p>Instituições como a American Psychological Association (APA), American Educational Research Association (AERA) e o National Council on Measurement in Education (NCME) distinguem quatro tipos fundamentais de validade:</p> <ol style="list-style-type: none"> <li>1 - conteúdo,</li> <li>2 - concorrente,</li> <li>3 - preditiva, e</li> <li>4 - construto</li> </ol>
<p>Validade de conteúdo – um escore é interpretado como refletindo a capacidade do examinando em relação a certos conteúdos e comportamentos.</p>
<p>A validade de conteúdo – também chamada de validade curricular, amostral ou lógica – é, dentre os diferentes tipos de validade, a que mais interessa aos testes de escolaridade.</p>
<p>Um teste é considerado como tendo validade de conteúdo quando constitui uma amostra representativa de conhecimentos e comportamentos adquiridos durante o processo educacional.</p>
<p>A validade de conteúdo refere-se não apenas à representatividade do conteúdo lecionado, mas também à das habilidades desenvolvidas.</p>
<p>A validade de conteúdo não é determinada estatisticamente, não é expressa por um coeficiente de correlação, mas sim resulta do julgamento de diferentes examinadores, que analisam a representatividade dos itens em relação às áreas de conteúdo e à relevância dos objetivos a medir.</p>
<p>O planejamento do teste tem grande influência na validade de conteúdo, pois é nesse momento que se organiza uma amostra representativa de conhecimentos e de comportamentos.</p>
<p>O problema central, na validade de conteúdo, é, portanto, o da representatividade da amostra, que, nos testes de escolaridade, é controlada através de tabelas de especificação.</p>
<p>O importante no teste, e que garante a validade de conteúdo, é a alta correlação entre o item e os objetivos do curso.</p>
<p>Há necessidade de estabelecer um equilíbrio entre o nível da matéria (assunto) e a capacidade do examinando. Alguns itens são corretos quanto à técnica de construção, mas não são válidos, porque testam capacidades que o examinando não possui ou pretendem verificar comportamentos que já são do conhecimento do examinador.</p>
<p>Validade concorrente – um escore é interpretado como representando uma característica não diretamente medida pelo teste.</p>
<p>A validade concorrente - também chamada de validade de status - é a relação entre os escores de um teste e os escores de um critério, obtido simultaneamente com os primeiros, e que medem o mesmo desempenho.</p>
<p>Um teste informa apenas sobre aquilo que está medindo e nada mais; entretanto, algumas vezes, procura-se saber algo mais, por exemplo, se há relação entre os escores do teste e um critério contemporâneo do desempenho na variável que o teste pretende medir, isto é, se o teste</p>

<p>possui validade concorrente. Assim, se um sujeito se saiu bem num teste de ortografia e, a seguir, fez uma dissertação, é de supor que também se saiu bem na dissertação e não praticou erros de grafia que também não cometeu no teste objetivo; nesse caso, o teste possui validade concorrente.</p>
<p>O elemento tempo é a principal diferença entre a validade concorrente e a validade preditiva. A simultaneidade da obtenção dos escores do teste e dos escores do critério identifica a validade concorrente.</p>
<p>Validade preditiva – um escore é usado para prever um comportamento no futuro.</p>
<p>Também chamada de validade empírica ou estatística - é o grau de correlação entre os escores de um teste e outras medidas do desempenho (critério), obtidas independentemente do primeiro teste. Se a correlação entre os escores do teste preditor e os escores da variável critério é alta, diz-se que o teste é válido para o fim a que se destina e que possui validade preditiva.</p>
<p>Os escores da variável critério não necessitam ser, forçosamente, os escores de um teste, podem ser classificações, aprovação ou reprovação num curso, sucesso ou insucesso numa atividade, erros cometidos numa tarefa etc..</p>
<p>O elemento básico no estabelecimento da validade preditiva é o critério. Sem um critério perfeitamente definido nas suas características, que seja fidedigno e que também seja válido, é impossível dizer o que o teste efetivamente mede, pois suas características são definidas pelo critério.</p>
<p>O coeficiente de validade permite prever, aproximadamente, o escore que o examinando possivelmente obterá na variável critério, com base no escore obtido num teste preditor.</p>
<p>O problema da validade é de grande relevância na área educacional (Vianna, 1983), tendo em vista o fato de que a avaliação se vale, frequentemente, de construtos, que, após sua definição operacional, são medidos por intermédio de testes.</p>
<p>Construto são traços, aptidões ou características supostamente existentes e abstraídos de uma variedade de comportamentos que tenham significado educacional (ou psicológico). Assim, fluência verbal, rendimento escolar, aptidão mecânica, inteligência, motivação, agressividade, entre outros, são construtos.</p>
<p>A validade de conteúdo e a validade de critério (concorrente e preditiva), não se preocupam, entretanto, com a compreensão dos construtos que os testes medem; conseqüentemente, impõe-se, uma nova abordagem para análise dos instrumentos de medida da aprendizagem escolar.</p>
<p>A validade de construto possibilita determinar qual a característica educacional que explica a variância do teste, ou, então qual o significado do teste. Ou, ainda, procura explicar as diferenças individuais nos escores desse instrumento de medida. O interesse, na validação de construtos, centraliza-se na característica ou traço que está sendo</p>

medido, mais do que no próprio teste.
A validade de construto é de particular importância sempre que um instrumento deva ser interpretado como proporcionando medidas de um atributo ou qualidade que, presumivelmente, as pessoas possuem.
A validade de construto, ao contrário da validade empírica, não é expressa em termos de um coeficiente quantitativo, conforme ocorre no caso da validade preditiva e da validade concorrente.
O conceito de validade de construto, por sua vez, é extremamente útil para explicar a natureza dos instrumentos que medem traços para os quais não se possuem critérios externos. Assim sendo, é necessário partir de uma variável logicamente definida. A variável, como um construto lógico, é inserida num sistema de conceitos cujas relações são explicadas por uma teoria e a partir da qual certas consequências práticas, sob determinadas condições, podem ser extraídas e testadas. Se o resultado é o que se esperava em uma série de testes, o instrumento é considerado como possuindo validade de construto para a variável testada.
É comum em educação o desenvolvimento de sistemas unificados de princípios, definições, postulados e observações, para explicar o relacionamento entre variáveis, ou seja, o desenvolvimento de teorias educacionais, a partir das quais são construídos instrumentos para a mensuração de um determinado traço ou característica. Somente por intermédio da constatação da validade de construto desses instrumentos desses instrumentos é que se pode confirmar o significado dessas características ou traços apresentados pela teoria. E para alcançar esse objetivo, necessário se faz a aplicação dos procedimentos clássicos do método dedutivo: teoria, dedução, hipótese, experimentação e, finalmente, dados que confirmem ou neguem a hipótese, ou seja, o construto.
A lógica da validade de construto, assim como o seu processo, é essencialmente, a do método científico. Parte-se de uma teoria sobre a natureza do construto e fazem-se predições sobre as relações entre os escores do teste e outras variáveis. A seguir, essas predições são verificadas empiricamente e, dependendo dos resultados, a teoria é aceita, revista ou rejeitada. O trabalho contínuo de fazer predições, testar hipóteses, através da experimentação, e rever a teoria são atividades que contribuem para aumentar a precisão da definição do construto.
As teorias que procuram explicar um determinado fenômeno compreendem um conjunto inter-relacionado de conceitos, proposições e leis. A esse sistema interligado Cronbach e Meehl (1955) deram o nome de rede nomológica.
É condição necessária, na pesquisa da validade de construto, que a definição de conceitos ou formulação de leis estejam apoiados, direta ou indiretamente, em dados observáveis.
Na validação de um construto há necessidade de um grande número de diferentes operações, inclusive de ordem qualitativa, para a mensuração

de um conceito e a fim de mostrar que esse mesmo conceito está ligado a outro conceito por intermédio de uma rede nomológica.
<p>É preciso considerar as seguintes situações:</p> <ol style="list-style-type: none"> <li>1) O processo de inferência do significado de um construto exige que os dados sejam observáveis;</li> <li>2) O processo de inferência do significado de um construto, a partir de dados observáveis, deve ser explicitamente especificado, para que se possa verificar a precisão de uma inferência;</li> <li>3) Diferentes usuários de um construto devem servir-se, essencialmente, de uma mesma rede nomológica, para que a concordância entre os pesquisadores seja possível, pois, frequentemente, o nome de um mesmo construto é usado com referência a diferentes construtos ou diferentes nomes são empregados com relação ao mesmo construto.</li> </ol>
A validação de construtos não se limita apenas a validar um teste, o seu alcance é bem mais amplo, centrando-se o seu objetivo na validação da teoria em que se apoiou a construção do instrumento; desse modo, o trabalho de validação de um construto é uma pesquisa científica empírica, porque, definidos os construtos que seriam responsáveis pelo desempenho no teste, o avaliador passa a formular hipóteses sobre a teoria dos construtos e, a seguir, testa empiricamente essas hipóteses.
Validade, na perspectiva de Messick, é um julgamento avaliativo integrado, baseado em evidências empíricas e fundamentos teóricos, sobre a adequação e a propriedade (no sentido de pertinência) das inferências a partir de escores ou outros modos de avaliação.
<b>MORROW JUNIOR, J. R. et al. Medida e avaliação do desempenho humano. 4. ed. Porto Alegre: Artmed, 2013. 472 p.</b>
Um teste deve primeiro ser reproduzível para que seja válido - para que de fato meça o que se propõe a medir.
A validade também pode ser amplamente classificada como de natureza lógica ou estatística. Sempre existe um tipo de critério independentemente do tipo de procedimento de validação envolvido. O critério pode ser percebido como a medida mais verdadeira que você está tentando medir.
A validade de conteúdo é uma evidência de veracidade baseada em tomada de decisão e interpretação lógicas. Os termos validade e validade lógica são frequentemente usados para a validade de conteúdo. O universo de interesse ou o universo de conteúdo para determinado teste precisa ser bem definido.
O fato de um teste refletir a validade do conteúdo, no entanto, não o torna necessariamente válido.
O critério para a validade de conteúdo existe na mente do interpretador. Os especialistas de conteúdo juízes, colegas e autores de livros didáticos podem servir como fontes para a validação de conteúdo de instrumentos. Os professores que desenvolvem testes cognitivos elaboram itens escritos que refletem o conteúdo do curso (então os itens são conteúdos válidos).

<p>A validade de critério é baseada na disponibilidade de uma medida de critério verdadeira. A validade tem por base a determinação da relação sistemática entre o critério e outras medidas usadas para estimar o critério. Em resumo, a validade de critério é a evidência de que um teste tem relação estatística com a característica que está sendo medida.</p>
<p>Outros termos para validade de critério são <i>validade estatística</i> e <i>correlacional</i>; esses termos são usados em decorrência de a evidência relacionada ao critério ser baseada no coeficiente de correlação linear de Pearson entre determinado teste e o critério.</p>
<p>... a evidência relacionada ao critério é muitas vezes subdividida em validade corrente e validade preditiva. Ambas são baseadas no coeficiente de correlação linear de Pearson.</p>
<p>A principal diferença entre elas é o momento em que o critério é medido. Para a validade concorrente, o critério é mensurado aproximadamente ao mesmo tempo que a medida alternativa. O critério é medido "no futuro" com a validade preditiva. Para estabelecer a validade preditiva, o critério pode ser avaliado em muitas semanas, meses ou mesmo anos depois de o teste original ter sido conduzido.</p>
<p>A validade de construto é muitas vezes usada para validar medidas que não são observáveis, mas que existem teoricamente.</p>
<p>A validade de construto é, essencialmente, um casamento entre a lógica (conteúdo) e os procedimentos de validade de estatística. Para fornecer a evidência da validade de construto para determinada medida, deve-se reunir uma variedade de informações estatísticas que, quando vistas em conjunto, adicionam evidências à existência do construto teórico que está sendo medido.</p>
<p>O trabalho de hipótese poderia ser desenvolvido da seguinte maneira quando se está coletando a evidência de validade de construto: se, na teoria, o construto é válido, então tais e tais coisas deveriam ocorrer. Em seguida, experimente para ver se realmente acontecem.</p>
<p>A parte lógica da validade de construto refere-se ao que poderia ocorrer. A parte estatística consiste nos dados coletados.</p>
<p>A coleta contínua de informação que confirme a teoria adiciona a evidência para a existência do construto. Quando aquilo que deveria ocorrer não é confirmado pela coleção de dados, existem dois aspectos a serem considerados: pode ser que o construto não exista ou que a relação lógica (se x, então y) tenha sido imprecisa.</p>
<p>O desenvolvimento da validade de construto é altamente relacionado ao método científico. Uma hipótese é gerada, um método é desenvolvido, os dados são coletados e analisados, e uma decisão é tomada com base na evidência obtida.</p>
<p>Muitas vezes, você ouvirá as palavras evidência "convergente" ou "discriminante". Em geral, esses conceitos são baseados nas correlações entre as variáveis. As variáveis que teoricamente deveriam se correlacionar fornecem provas de evidência convergente, e as variáveis que, em teoria, não deveriam ter correlação fornecem provas discriminantes (i. e., elas medem aspectos diferentes).</p>

**BLACK, Paul. Os professores podem usar a avaliação para melhorar o ensino? Tradução de Fernando Zan Vieira. Práxis Educativa, Ponta Grossa, v.4, n.2, jul.-dez. 2009. Disponível em: <<http://www.periodicos.uepg.br>>. Acesso em: 15 jun.2017. Tradução de: Can teachers use assessment to improve learning?**

Testes externos gozam de maior confiança pública do que deveriam. Os limites de sua validade e confiabilidade não são compreendidos. Um conceito amplo de validade é ao mesmo tempo necessário e persuasivo em qualquer tentativa de melhorar a qualidade das avaliações. A afirmação inicial da revisão feita por Messick (1989) deu uma definição legítima: "validade é um julgamento avaliativo integrado do grau no qual a evidência empírica e as justificativas teóricas fundamentam a adequação e a propriedade de inferências e ações baseadas nos escores de testes ou outros modos de avaliação."

Validade – característica que torna um procedimento ou um instrumento aceitável, e que faz com que, graças a isso, se possa ter confiança nos resultados.

Um instrumento é válido quando consegue medir o que é pressuposto medir, e leva a um resultado quando exprime o que tem por função exprimir.