

# Como avaliar práticas docentes?

Instrumentos e evidências empíricas

**Dados Internacionais de Catalogação na Publicação (CIP)  
(Câmara Brasileira do Livro, SP, Brasil)**

Como avaliar práticas docentes [livro eletrônico] :  
instrumentos e evidências empíricas / [Flávio  
Comim ; coordenação Esmeralda Correa Macana]. --  
São Paulo : Fundação Itaú para a Educação e  
Cultura, 2022.  
PDF

ISBN 978-65-86771-18-3

1. Avaliação educacional 2. Prática de ensino  
3. Professores - Formação profissional I. Comim,  
Flávio. II. Macana, Esmeralda Correa.

22-123452

CDD-370.71

**Índices para catálogo sistemático:**

1. Avaliação educacional : Prática docente :  
Educação 370.71

Cibele Maria Dias - Bibliotecária - CRB-8/9427

## **FUNDAÇÃO ITAÚ**

### **Conselho Curador**

#### **Presidente**

Alfredo Egydio Setubal

#### **Vice-presidentes**

Ana Lúcia de Mattos Barreto Villela

Maria Alice Setubal

#### **Conselheiros**

Claudia Politanski

Danilo Santos Miranda

Eduardo Queiroz Tracanella

Heitor Sant'anna Martins

Osvaldo do Nascimento

Priscila Fonseca da Cruz

Ricardo Manuel dos Santos Henriques

Rodolfo Villela Marino

### **Diretoria**

#### **Presidente da Fundação Itaú**

Eduardo Saron Nunes

#### **Diretores**

Álvaro Felipe Rizzi Rodrigues

Cristiano Angulski de Lacerda

Luciana Nicola Schneider

Paulo Sergio Miron

Valéria Aparecida Marreto

### **Itaú Social**

#### **Gerente de Pesquisa e Desenvolvimento**

Patrícia Mota Guedes

#### **Coordenadora da Pesquisa**

Esmeralda Correa Macana

#### **Autoria**

Flávio Comim

#### **Assistentes de pesquisa – Universidade de Cambridge**

Fernando Bucheli

Caroline Souza

Jonathan Lynn

Marcela Moraes

Ningshan Zhang

Charne Thomson

### **Comunicação Integrada Fundação Itaú**

#### **Gerente de Comunicação**

Ana de Fátima Oliveira de Sousa

#### **Coordenação de Comunicação para a Educação**

Alan Albuquerque R. Correia

#### **Equipe de Comunicação**

Tayrine Mauricio

## Índice Geral

Resumo Executivo	06
Introdução	09
1. O que é um professor de qualidade?	11
2. Como avaliar professores?	14
2.1 Lições do Projeto MET	15
2.2. Aprofundando os diferentes instrumentos de Avaliação	19
2.3 Comparação entre os diferentes instrumentos	38
3. Critérios técnicos para Avaliação de testes	43
4. Implicações Práticas (para políticas públicas)	46
5. Consulta Empírica	48
6. Lições das Experiências Internacionais	52
7. Conclusões	58
8. Referências	60
9. Anexo com instrumentos	78

## Índice de Tabelas, Quadros e Diagramas

Diagrama 1 – A centralidade da questão avaliativa	10
Tabela 1 – Exemplos de atributos críticos nas rubricas FFT	21
Diagrama 2 – Estrutura Analítica do CLASS	22
Diagrama 3 – Critérios para rubricas	23
Diagrama 4 – Elementos do PLATO	24
Tabela 2 – Qualidades dos Professores efetivos	28
Diagrama 5 – Etapas do INFORM	30
Diagrama 6 – Modelo Reflexivo	34
Diagrama 7 – Evolução dos Modelos de auto eficácia	36
Tabela 3 – Instrumentos de observação, visão pedagógica e dimensões	38
Tabela 4 – Comparação entre instrumentos de avaliação e dimensões	41
Tabela 5 – Prós e contra de cada método	42
Quadro 1 – Desenho das observações calculando os vieses	45
Diagrama 8 – O que é um professor de qualidade?	49
Diagrama 9 – Ranking de fatores que afetam o desempenho dos professores	50
Tabela 6 – Utilidade dos modelos de avaliação	51
Tabela 7 – Experiências Internacionais de Avaliação de professores	52
Diagrama 10 – Tipos de avaliação de professores no mundo	57

## Resumo Executivo

O objetivo desse estudo foi o de catalogar, sistematizar e comparar diferentes instrumentos e métodos de avaliação de professores. Identificamos 12 classes de modelos principais de avaliação de professores na literatura especializada (modelos de valor-adicionado, modelos de observação em classe, surveys de opinião de estudantes, surveys de opinião de professores, surveys de opinião de diretores, modelo INFORM de observações de campo, análise de portfólio dos professores, modelos de simulação, modelos de práticas reflexivas, análise do conhecimento dos professores, modelos de auto eficácia e modelos mistos).

Vimos que os métodos mais usados não são necessariamente os mais abrangentes e que a melhor estratégia é a de produção de indicadores e metodologias mistas, como propostas pelo projeto MET da Fundação Gates. Não existe uma fórmula mágica para a geração desses indicadores. No entanto, descobrimos que é possível melhorar essas metodologias mistas pela inclusão de indicadores não contemplados nos estudos da Fundação Gates, como indicadores de auto eficácia ou de práticas reflexivas envolvendo 'debriefing'.

Em relação ao uso potencial dos resultados das avaliações vimos que a lição principal foi a de que seu uso depende prioritariamente do tipo de medida utilizada. Instrumentos de percepção ou de observação de classe não têm o mesmo potencial sistemático de medidas de valor-agregado, que por sua vez não têm o potencial de serem usadas para formação de professores. Talvez não existisse tanto objeção à avaliação de professores, como encontramos em alguns países, se seus usos fossem melhor especificados e esclarecidos. Se positivas, as avaliações podem ser usadas para incentivos, promoção e alocação de professores. Se negativas, podem ser usadas para preparar melhor os piores professores ou no limite demiti-los, o que pode naturalmente causar muita apreensão e incerteza. Por essa razão, objetivos claros a priori são fundamentais para a implementação de processos de avaliação de professores.

Identificamos 3 critérios principais largamente utilizados para avaliar os diferentes instrumentos de avaliação de professores, quais sejam, viés, confiabilidade e validade. No entanto, encontramos também critérios menos utilizados como justiça, 'accountability' e estabilidade.

Diferentes modelos de avaliação, baseados em distintos modelos pedagógicos, têm implicações diferenciadas sobre os melhores modos de capacitar os professores. Em linhas gerais é possível sugerir um conjunto de práticas derivadas da discussão acima, que dever tornar mais eficiente a capacitação dos professores, tais como:

- 1.** Professores devem ser capacitados em gestão de classe. Não basta transferir conhecimento ou informação aos professores. Eles devem ser capazes de internalizar conhecimentos práticos, concretos que os ajudem a navegar melhor entre seus alunos e contextos difíceis;
- 2.** Professores devem ser capacitados em atributos psicoemocionais que são indispensáveis, segundo as evidências encontradas, para um ensino eficiente;
- 3.** Professores devem ser capacitados através de instruções concretas que possam ser entendidas como parte do seu dia-a-dia; capacitações em teorias abstratas não tem grande potencial de mudança do comportamento concreto dos professores;
- 4.** Professores devem ser capacitados para lidar com o imprevisto e o inesperado, pois isso é um padrão identificado como vigente em muitas salas-de-aula;
- 5.** Pela mesma medida, professores são melhor capacitados dentro do seu ambiente de trabalho (escolas) ao invés de o serem em outros ambientes que aumentem a dificuldade de transferência do seu conhecimento (pois esse deve ser prático, levando em consideração as circunstâncias de aplicação)
- 6.** Capacitações devem respeitar o tempo de interpretação, apropriação e transformação dos professores; não se deve esperar que os professores sejam apenas reprodutores de fórmulas pedagógicas definidas em gabinetes. Muitas vezes isso significa também permitir um tempo de apropriação coletiva;

**7.** As evidências coletadas sugerem que modelos mistos de avaliação parecem ser os mais frutíferos pois permitem o estabelecimento de parâmetros de validade e confiabilidade que atendam diferentes propósitos. Se possível, feitos em consulta com os próprios professores e atendendo os protocolos específicos de cada instrumento específico.

Encontramos que a experiência internacional de uso de modelos e instrumentos de avaliação de professores é muito heterogênea. Os países que avaliam seus professores diferem no grau de centralização/descentralização da aplicação de exames padronizados, no grau de autonomia que conferem a gestão e à avaliação nas escolas, na periodicidade das avaliações, na existência ou não de departamentos especializados na implementação das avaliações, no uso de instrumentos diferenciados, no link com a formação dos professores e uso ou não de medidas de certificação, e finalmente no estágio curricular em que implementam essas avaliações.

Como parte desse estudo foi realizada na metade de 2018 uma pesquisa empírica sobre o tema da avaliação de professores com gestores educacionais, gestores públicos, professores do ensino fundamental e universitário, especialistas em educação, jornalistas e pesquisadores da área educacional. Foram entrevistadas 236 pessoas com uma amostra experimental. Os seguintes resultados se destacaram:

- 1.** Para as pessoas consultadas 63% dos fatores que definem um bom professor estão relacionados a seus aspectos psicoemocionais (como comprometimento, amor, dedicação, vontade, etc)
- 2.** Somente 15% das pessoas consultadas associam um professor auto eficaz a uma boa gestão de sala-de-aula
- 3.** As pessoas consultadas consideram “a influência dos professores” o maior fator determinante das notas dos alunos, seguido da “influência das famílias” e por último da “influência da direção”
- 4.** Parece haver um desacordo entre os entrevistados sobre o grau desejado de intervenção dos professores em sala-de-aula. De um lado encontramos 42.7% que acham que os



professores devem intervir em tudo que podem em classe; de outro 49.6% acreditam que os professores devem tentar resolver tudo pelo diálogo. Enquanto isso 7.7% revelou pensar que os professores não devem intervir em nada e deixar a aula rolar segundo a vontade dos alunos.

**5.** 85.4% das pessoas acham que os professores do ensino médio deveriam ser avaliados em bases regulares. É relevante notar que não há diferença significativa na visão de gestores e professores do ensino fundamental nesse quesito. Para 64% dos entrevistados a maior vantagem nesse tipo de avaliação é poder contar com uma medida de melhoria de qualidade do ensino. Um percentual de 12% acha que as avaliações podem ser utilizadas como medida de promoção (incentivos) aos professores. Por outro lado, 10% acha que as avaliações possam ter nenhuma utilidade. Um mesmo percentual de 10% acredita que as avaliações podem ser usadas para a promoção de capacitação daqueles professores menos preparados.

**6.** Vemos que os modelos de valor-adicionado, surveys de opinião de estudantes, professores e diretores estão entre os considerados 'nada úteis' enquanto modelos voltados à observação de classe, análise de portfólio, conhecimento e práticas reflexivas de professores estão entre os considerados mais úteis. Nenhum, no entanto ganha dos chamados modelos mistos (onde vários desses modelos são combinados

**7.** Um alto percentual, de 80.5% das pessoas entrevistadas, considera que algum desses modelos é muito útil. Combinando essa estatística com a anterior sobre a utilidade das avaliações regulares podemos ver, para esse conjunto de pessoas entrevistadas, que há um ambiente favorável as discussões sobre a melhor forma de avaliação dos professores.

Esse trabalho catalogou uma série de instrumentos incluídos no anexo que podem ser úteis na elaboração de instrumentos de avaliação de professores por aquelas pessoas interessadas.

## 1. Introdução

Existem poucos assuntos tão debatidos em todo o mundo como a educação. De fato, especialistas de disciplinas tão distintas como pedagogia, filosofia, economia, sociologia, antropologia, psicologia e neurociência (para ficar apenas em algumas) têm todos algo a dizer sobre a educação a partir de uma fórmula diferente, um ângulo divergente no qual olham o processo de aprendizagem e desenvolvimento de nossas crianças e jovens. O debate *per se* é inevitável (e saudável). Apesar das diferenças de contexto e linguagem, chama atenção que quase todas essas perspectivas distintas coincidem em um fator fundamental: a importância chave e estratégica do professor para a educação. Dizer isso parece uma redundância, mas dado o conjunto de teorias e abordagens diversas para a questão educativa, é importante estabelecer com clareza esse ponto de partida.

No nosso caso particular, tratamos não apenas das vidas dos quase 2,2 milhões de docentes que atuam na educação básica brasileira (onde aproximadamente  $\frac{3}{4}$  encontram-se na rede pública) mas de maneira direta de quase 49 milhões de crianças e suas famílias. De maneira indireta, levando em conta os efeitos sobre a formação de capital humano no país e seus impactos sobre a sociedade e economia, falamos de números ainda mais expressivos. Mesmo assim, apesar do reconhecimento, da essencialidade e do alcance do trabalho do professor, muito pouco ainda é estudado à luz das evidências nacionais e internacionais sobre como melhor avaliar o trabalho dos professores. Mais especificamente, se um bom professor é um ingrediente tão importante para a educação (senão o 'mais importante') por que não usamos o que sabemos sobre como avalia-los e por quê não temos estabelecidos mais protocolos de avaliação de professores no Brasil? Uma primeira resposta para problemas que poderiam ser de implementação poderia ser a falta de mais estudos que sistematizem evidências sobre instrumentos e modelos de avaliação de professores.

Estudos realizados sobre avaliações feitas com professores, como o de Jacob e Lefgren (2008) ou a série de estudos feitos pela fundação Bill e Melinda Gates (2013), coincidem que essa não é uma tarefa trivial e que há uma grande heterogeneidade na qualidade dos professores que não é simplesmente explicada por suas características observáveis, tais como experiência, educação ou tipo de contrato. Por outro lado, existe um histórico feito de avaliações onde um número muito pequeno de professores é identificado como

insatisfatório, sugerindo que os sistemas de avaliação atuais são pouco úteis para a gestão escolar por não oferecer instrumentos de discriminação entre eles. E parece também serem pouco úteis aos professores pois esses relatam receber pouco ou quase nenhum retorno dos sistemas atuais de avaliação.

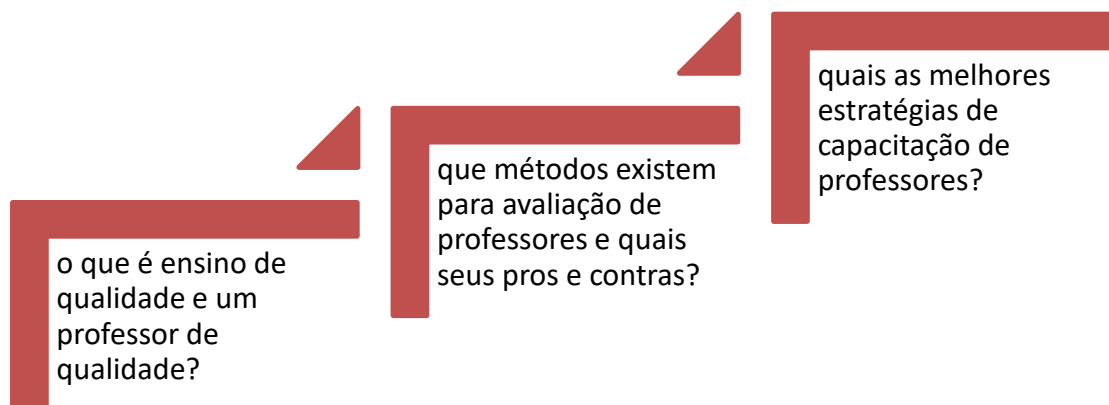
Portanto, se para avançar na melhoria da educação precisamos melhores professores e se para isso é importante que saibamos oferecer a eles e elas melhores formações, então é fundamental que tenhamos elementos provenientes de avaliações para saber o que dá melhores resultados com os alunos, entendido aqui resultados não somente como parâmetros de desempenho provenientes de testes padronizados, mas de uma formação ampla de habilidades cognitivas e não-cognitivas dos alunos.

Nesse contexto, o objetivo desse trabalho é oferecer uma compilação de diferentes instrumentos que existem para avaliação de professores e examina-las à luz das evidências disponíveis. Procura-se sistematizar e comparar esses instrumentos ao mesmo tempo em que se oferece uma análise crítica. Essa ênfase no estudo de diversos instrumentos e metodologias é justificada por um consenso na literatura acerca da ineficácia de métodos convencionais de avaliação de professores e pela necessidade de reformas pedagógicas.

Em grandes linhas esse trabalho articula três literaturas que aparecem convencionalmente separadas, mas que representam uma linha de continuidade lógica necessária à melhoria da formação dos professores. A primeira linha examina aquilo que se quer avaliar, ou seja, o que é um professor/a e um ensino de qualidade. Sem essa referência normativa parece fazer pouco sentido o debate seguinte sobre avaliação de professores. É óbvio que do ponto de vista técnico diferentes métodos podem ser usados para avaliar diferentes características desejáveis dos professores, mas há um elo importante entre métodos e normativas. Por exemplo, uma normativa que identifique autonomia e cooperação como características importantes no ensino de qualidade não pode ser bem avaliada unicamente com métodos e instrumentos totalmente individuais porque deixam de considerar parte do fenômeno, nesse caso de interação, que não pode ser totalmente captado a partir de aspectos exclusivamente individuais. Por fim, há também um elo entre métodos de avaliação e estratégias de capacitação de professores. É importante, contudo destacar que a existência dessas relações não diminui em nada a centralidade da questão da avaliação dos professores. Pelo contrário, mostra como ela é fundamental para articular outros

campos de discussão. Esquemáticamente podemos representar essas relações no diagrama abaixo:

### Diagrama 1 – a centralidade da questão avaliativa



Esse trabalho mapeia diferentes níveis de complexidade nos quais se pode entender a qualidade de um professor, indo do mais simples, vinculado ao conhecimento curricular dos mesmos até os mais sofisticados que tratam das habilidades dos professores em lidar com situações inesperadas dentro de sala-de-aula. Logo depois, oferece um inventário de diferentes instrumentos e metodologias de avaliação procurando compara-las e avalia-las dentro de uma perspectiva de desenvolvimento humano. Como complemento dessa discussão esse trabalho avalia os resultados de uma pesquisa empírica feita com 236 *stakeholders* diretos da educação no Brasil, na sua maioria gestores/as educacionais. Na sequência, examina evidências provenientes de 31 países sobre como estes avaliam seus professores, para concluir com algumas sugestões e recomendações sobre como melhor avaliar professores e como ligar isso com políticas de formação de professores.

#### 1. O que é um professor de qualidade?

Existem cinco grandes blocos de critérios que definem um professor de qualidade na literatura examinada. O primeiro, que não será tratado aqui, vincula a qualidade de um professor à obediência a um certo paradigma pedagógico particular. Como existem diversas tradições pedagógicas no mundo há uma certa irredutibilidade e incomensurabilidade

natural nesse primeiro critério, o que o torna menos útil. Em outras palavras, diferenças paradigmáticas parecem pouco produtivas como critérios de atribuição de qualidade dos professores uma vez que existem diversos paradigmas pedagógicos não comparáveis. É certo também que esse critério parece respeitar pouco a independência e diversidade dos professores e de seus contextos educacionais.

Um segundo conjunto de critérios, como discutido por Belsito (2016), destaca as diferenças em avaliar um 'ensino de qualidade' vs um 'professor de qualidade'. Por *ensino de qualidade* entende-se o resultado final do processo educativo que é em parte produzido pelo professor, mas que também depende de fatores que vão além dos professores, como a gestão escolar e modelos pedagógicos. Por outro lado, quando falamos unicamente da *qualidade do professor* tratamos daqueles fatores que são exclusivamente atribuíveis aos professores, tais como suas atitudes, suas ações de construção de relacionamentos com os estudantes, suas expectativas em relação aos estudantes, seus métodos utilizados, estratégias e ações implementadas em classe e não menos importante suas capacidades de gestão de sala de aula. A qualidade de um professor inclui muitos aspectos intangíveis e complexos que não podem ser capturados por fatores tais como qualificações e testes de habilidades acadêmicas, como reconhecido pela OCDE (2005).

Em outras palavras, o conceito de *ensino de qualidade* comporta a resultante de todos aqueles fatores que fazem a vivência e aprendizado dos alunos e alunas terem qualidade enquanto o conceito de *professor de qualidade* trata apenas daquelas características individuais e pessoais dos professores, resultem elas ou não em um ensino de qualidade. Claro, esperamos que um professor de qualidade seja um elemento decisivo para um ensino de qualidade. Mas pode haver a possibilidade de que essa contribuição dos professores seja compensada por fatores não diretamente relacionados aos mesmos, como aspectos de gestão da escola, modelos pedagógicos utilizados, falta de recursos, etc.

Grande parte da literatura consultada (como por exemplo Abu-Tineh, Khasawneh e Khalaileh, 2011, Yilmaz e Cavas, 2008) enfatiza a gestão de sala-de-aula como o elemento chave na explicação da qualidade dos professores. A questão pode ser mais complexa pois podem existir diferentes abordagens usadas na gestão de sala-de-aula como a i) intervencionista, ii) não-intervencionista e iii) interativa, indo de um alto grau de controle até uma divisão de responsabilidades. O que parece ser essencial é que independente da

abordagem utilizada deve haver uma crença por parte do professor na sua habilidade em realizar o trabalho desejado.

Um terceiro conjunto de critérios (que possui algum overlap com o conjunto acima) se fundamenta no conceito de 'auto eficácia' (do inglês, "self-efficacy") dos professores. Um professor de qualidade seria aquele que acreditaria que ele ou ela teria a capacidade para influenciar o desempenho de seus estudantes, mesmo aqueles que fossem difíceis de ensinar. A chave da discussão dessa literatura, que se iniciou com a teoria do aprendizado social de Rotter (1966) e que amadureceu com os trabalhos da Corporação RAND de 1976 e Bandura (1977), é como se distingue aqueles fatores que os professores podem controlar daqueles que não podem. No trabalho clássico de Tschannen-Moran, Woolfolk-Hoy e Hoy (1998) eles destacam que a auto-eficácia é um processo cognitivo no qual os professores constroem crenças sobre sua capacidade de ter um impacto sobre os seus alunos/as que de fato determina quanto esforço eles/as colocam nas suas aulas e condicionam como eles persistem em relação aos obstáculos que encontram assim como lidam com os fracassos que possam ocorrer.

Esse conceito é interessante pois contempla a necessidade de que os professores façam suas próprias avaliações de suas habilidades pedagógicas. Nesse contexto, um método de avaliação *top-down* não teria como produzir os efeitos desejados. Seria mais apropriado utilizar instrumentos como o TSES (do inglês, "Teacher Self-Efficacy Scale") ou o ABCC (do inglês, "Attitudes and Beliefs on Classroom Control"). Enquanto o TSES foca em duas dimensões principais, chamadas de eficácia pessoal do professor e eficácia geral do professor (essa última leva em consideração a capacidade do professor em contrapor-se as influências negativas do background dos estudantes) medidas através de 22 itens com uma escala Likert de 4 pontos, o ABCC está estruturado em 3 dimensões, gestão da informação (14 itens), gestão das pessoas (8 itens) e gestão do comportamento (4 itens).

Um dos desafios dessa literatura é separar a influência das famílias no desempenho dos alunos da influência das escolas. A evidência internacional, como discutida por Ko & Sammons (2012), sugere que o impacto máximo dos professores no desempenho dos alunos (em escolas primárias onde frequentemente as crianças passam o dia todo com o professor) é da ordem de 40%. No entanto, cabe notar que a influência potencial das

escolas é maior entre os grupos menos privilegiados e socialmente excluídos (Mincu, 2015). Assim, a qualidade dos professores é essencial para que estudantes pobres possam ganhar algo com as escolas. Há um consenso relevante de que se atinge uma melhoria de ensino mais consistentemente através do fortalecimento da capacidade dos professores ao invés de se reforçar controles sobre esses.

Uma quarta literatura se organiza a partir de uma distinção teórica entre ‘conhecimento dos conteúdos’ vs ‘conhecimento dos conteúdos pedagógicos’ (no inglês se usa as siglas CK para “content knowledge” e PCK para “pedagogical content knowledge”) como por exemplo em Hill, Rowan e Ball (2005) e Kleickmann et al (2013). O ponto principal aqui não é de todo distante do discutido acima: um bom professor precisa não somente de conhecimento dos conteúdos que vai ensinar mas sim de um conhecimento pedagógico, de natureza prática, especializada, que contemple por exemplo a leitura do que acontece em sala-de-aula com seus estudantes e seus estados motivacionais (Sonmark et al, 2017). Isso não quer dizer, no entanto, que não existam restrições sobre o conhecimento limitado de professores em algumas áreas, como é o caso das ciências (Nilsson e Loughran, 2012). Por essa razão é importante que consideremos instrumentos como o *CoRe* (“Content Representations”) que podem representar e mapear como podem ser trabalhados conteúdos de modo concreto.

A questão da qualidade dos professores tem sido muito debatida pela OCDE, preocupada com o envelhecimento do corpo docente nos países membros e com altas taxas de atrito (“attrition”) entre professores novos, o que parece não ser um fenômeno exclusivamente brasileiro. Nesse contexto, instrumentos como o *Survey de Conhecimento dos Professores* feito dentro do projeto Aprendizado Efetivo para o Ensino Inovador - ITTEL TKS (do inglês, “Innovative Teaching for Effective Learning – Teaching Knowledge Survey”) são mais adequados como medidas de práticas em classe e uso de procedimentos de instrução.

Há um quinto e último conjunto de critérios derivados de um conceito de qualidade de professores baseados nas suas habilidades de promoção da criatividade e gestão de risco em sala-de-aula. A ideia é que um professor de qualidade é aquele que promove um ambiente de criatividade na sala-de-aula no qual eventos não-esperados podem acontecer. Rowland & Turner (2007) introduzem o conceito de ‘contingência’ para fazer sentido de como os professores respondem aos eventos de classe que não são antecipados pelos planos de estudo. No contexto internacional isso se refere principalmente a uma

observação ou pergunta que pode ser feita inesperadamente por um estudante. Mas em um contexto nacional isso poderia ser adaptado a contingências comportamentais ou problemas de disciplina ou violência em classe. Seja como for, um professor de qualidade seria aquele que apresentaria contingência, ou seja, a habilidade de responder adequadamente a uma intervenção dos alunos que está fora de seus planos ou expectativas. Como colocam Coles e Scott (2015), se trata de desenvolver novos modos de lidar com a complexidade do espaço da sala-de-aula. Eles examinam a importância dessa habilidade no contexto de aulas de matemática. Eles trabalharam com o modelo de 4 estágios de Chapman e Heater (2010) de mudança da prática de professores mapeando i) a experiência de uma tensão cognitiva e emocional, ii) o enfrentamento dessa tensão, iii) a interpretação da tensão como um problema e iv) a busca de soluções. Assim, procuraram avaliar a capacidade dos professores de lidarem com a complexidade do espaço da sala-de-aula e de suas consequências não esperadas. A evidência relatada por Coles e Scott (2015) foi de natureza etnográfica, com detalhes da evolução de um aluno e de um professor ao longo de um ano acadêmico. No contexto das aulas de matemática o elemento chave foi o reconhecimento da existência de padrões entre os números pelo aluno, o qual foi possível pelo fato do professor subordinar sua estratégia de ensino as necessidades contextuais do aluno.

Indo direto ao ponto: falar em qualidade de ensino e do professor é tratar não somente do conhecimento como resultado final, mas como um processo que precisa ser construído e avaliado; é contemplar o ambiente de aprendizado e a relação entre alunos e professores e seu sentido de organização e pertencimento (Hughes e Chen, 2011; Liberante, 2012, McArdle, 2010); é fazer da sala-de-aula um espaço de criação onde o mais importante é subordinar o processo de ensino ao aprendizado (e não limitar o aprendizado aos planos de ensino). Há evidências recentes que alguns elementos não-cognitivos, como a chamada 'empatia epistêmica' tem um papel importante a exercer na qualidade do ensino (Jaber et al, 2018).

Antes de terminar cabe enfatizar que antes de mais nada as avaliações de professores estão logicamente subordinadas ao conceito de qualidade dos professores que se quer avaliar. Por definição, um professor de qualidade é aquele que satisfaz determinados critérios. Se estamos falando de i) adoção de modelo pedagógico, ii) gestão de sala-de-aula, iii) auto eficácia, iv) conhecimento pedagógico, v) capacidade de lidar com situações



inesperadas ou mesmo outras características mais sócio emocionais como afeto dispensado a seus alunos ou capacidade colaborativa com outros professores, estamos já pré-definindo os resultados de determinado modelo de avaliação.

Em outras palavras, um professor de qualidade é aquele ou aquela que segue um modelo pedagógico considerado apropriado, que faça uma boa gestão da sua sala-de-aula, que seja auto eficaz, com conhecimento pedagógico relevante e que esteja preparado para lidar com situações inesperadas. Possivelmente esses são critérios que no conjunto são impossíveis de serem cumpridos, mas que oferecem um norte ou referência normativa que deve ser utilizada nas avaliações de professores.

## **2. Como avaliar professores?**

Tecnicamente falando podemos separar a avaliação dos professores em quatro grupos, chamados de i) instrumentos de mensuração (“measurement tools”) que procuram avaliar o quão efetivos são os professores na realização de seus trabalhos, ii) instrumentos de desenvolvimento profissional (“professional development tools”) que podem oferecer informação direta aos professores possibilitando seu crescimento profissional e aumento de efetividade, iii) instrumentos de prática reflexiva (“reflexive practice”) que desenvolvem o potencial reflexivo e avaliativo dos professores para que de modo auto-sustentável possam ter uma cultura de avaliação e iv) modelos mistos, que procuram combinar as virtudes de diferentes modelos que captam diferentes aspectos das práticas dos professores. A grande questão sem dúvida reside nos critérios de avaliação e escolha entre os diferentes instrumentos disponíveis. Em geral, instrumentos específicos de avaliação traduzem visões particulares sobre o significado de um ensino de qualidade.

### **2.1 Lições do Projeto MET**

Não há dúvida de que o debate internacional tem sido muito influenciado pelas experiências Norte-americanas nesses últimos anos, e em particular devemos mencionar o influente MET (do inglês “Measuring Effective Teaching” Project, ou seja o “Projeto de Medir o Ensino Efetivo”). Suas publicações principais, como Kane et al (2010), Kane et al (2012), Ho e Kane (2013) e Kane, McCaffrey, Miller e Staiger (2013), patrocinadas pela Fundação Bill & Melinda Gates, seguem uma sequência cronológica de primeiros resultados, análise madura e discussão conclusiva. Há também toda uma subliteratura que segue os passos

do que se avançou nesse projeto. É importante assim que paremos para avaliar suas principais lições.

O Projeto MET, iniciado em 2009, postulou no seu primeiro relatório “Learning about Teaching” (em português, “Aprendendo sobre Ensinar”) suas três premissas fundamentais, quais sejam, primeiro, que sempre que possível a avaliação de um professor deve incluir os progressos de desempenho de seus estudantes; segundo, que quaisquer outros componentes ou métodos usados para avaliação devem estar associados a esses progressos e por fim, que a medida de avaliação utilizada deve incluir sugestões sobre aspectos específicos da prática dos professores que apoie o crescimento e desenvolvimento de seu trabalho. Podemos considerar essas premissas adequadas ou não, mas é importante observar que há quase sempre um conjunto de decisões metodológicas envolvidas em projetos de avaliação que são inevitáveis. Melhor quando elas são explicitadas, como feito pelo MET.

O projeto focou inicialmente em 5 conjuntos de medidas de qualidade dos professores (todas discutidas em detalhe na sequência de nossa narrativa). O primeiro conjunto de medidas foi elaborado a partir de *ganhos em testes padronizados* (como exames de disciplinas tais como inglês, matemática, biologia feitos por estados nos Estados Unidos) obtidos por estudantes ensinados por diferentes professores. O segundo conjunto focou em medidas de observação de classe complementadas com reflexões de professores (incluindo aqui as medidas CLASS, PLATO, MQI (em português, “Qualidade da Instrução Matemática”) e o QST (em português, “Qualidade do Ensino de Ciências”). O terceiro conjunto tratou do conteúdo de conhecimento pedagógico possuído pelos professores (por exemplo, se professores são capazes de identificar erros comuns no raciocínio dos estudantes antecipando-os e corrigindo-os). Os dois últimos conjuntos de medidas foram formados por percepções, tanto de estudantes (usando um survey chamado TRIPOD, discutido abaixo) como de professores (usando um survey formulado pelo “New Teacher Center”) sobre as condições de classe e de trabalho.

De modo muito resumido pode-se dizer que as principais evidências obtidas nessa primeira fase do projeto sugeriram que: i) medidas de valor-adicionado têm forte poder de previsão das notas dos alunos. Claro que o valor-adicionado de um professor é sujeito a flutuações

anuais e varia de classe-a-classe, mas essa volatilidade não afeta a confiança na medida; ii) professores que agregam valor tendem a promover um melhor entendimento conceitual dos alunos; iii) os professores têm um impacto maior nas notas de matemática do que nas de inglês ou de artes, e por fim, iv) as percepções dos estudantes dos pontos fracos e fortes dos professores são consistentes entre grupos diferentes de estudantes. Um resultado interessante obtido a partir do uso da medida TRIPOD, de percepção dos estudantes, mostra que um ambiente ordenado, no qual os estudantes respeitam seus professores, é um dos fatores mais fortemente correlacionados com o valor-adicionado do professor em matemática (Kane et al, 2010: 24). Esse primeiro estudo também fez testes com medidas combinadas (tais como valor-adicionado com percepção dos estudantes) com resultados positivos.

Na segunda fase do projeto, como pode ser visto em Kane et al (2012), a medida QST para ciências foi substituída pelo UTOP (em português, “Protocolo de Observação de Professores), desenvolvido por Michael Marder e Candace Walkington da Universidade de Texas-Austin. Também foram introduzidos critérios de observação de sala-de-aula como confiabilidade e escalabilidade (ou ‘transferabilidade’, no sentido de que instrumentos de observação de professores possam ser usados não somente por *experts* mas por um grupo maior de observadores). Grande parte das comparações e análises são possíveis devido a equivalência utilizada das medidas de valor-adicionado em ‘meses de escolaridade. Mais especificamente foi usado um fator de conversão de 0.25 desvios-padrão para cada 9 meses de escolaridade. As evidências obtidas por esse estudo a partir da gravação de 7.491 vídeos feitos com 1.333 professores de 6 distritos diferentes nos Estados Unidos com 44.500 estudantes, mostrou que:

- Todos os instrumentos avaliados foram associados positivamente com uma melhoria de notas dos alunos;
- A avaliação das aulas dos professores difere muito segundo quem observe e segundo qual aula seja observada. As evidências obtidas sugerem que um grau maior de confiabilidade (na faixa de 0.65) pode ser obtido para 80% das medidas utilizadas se tivermos uma aula avaliada quatro vezes por um observador diferente;
- Uma combinação de medidas melhora a confiabilidade e o poder preditivo da avaliação dos professores. Por exemplo, quando se ampliou a medida de

observação de classe com o uso conjunto da opinião dos estudantes, obteve-se uma diferença no ganho de desempenho entre o quartil superior e inferior de professores de matemática de 2.6 meses para 4.8 meses. As evidências também sugerem que a combinação de medidas confere um grau maior de confiabilidade e estabilidade às avaliações;

- As medidas combinadas fornecem um instrumento mais apto para captar o valor-adicionado dos professores em comparação com suas respectivas experiências (com benchmarks de 3 anos e 12 anos de experiência) ou qualificação (com e sem mestrado) dos professores. Por exemplo, enquanto 12 anos de experiência de um professor adicionariam 0.5 meses a mais de aprendizado matemático nas crianças, professores mais aptos avaliados por medidas combinadas adicionariam quase 8 meses em comparação com os menos aptos;
- Professores com melhor desempenho em medidas combinadas também têm melhores notas em outros parâmetros relacionados a surveys com estudantes, como por exemplo, seu vínculo emocional com o professor, grau de satisfação com as classes e maior nível de dedicação necessária nos cursos.

Lidas em conjunto essas evidências recomendam o uso de várias medidas em conjunto além do tradicional combo 'valor-adicionado+observações de classe'. As vantagens em termos de satisfação de critérios desejáveis incluem um maior poder preditivo, confiabilidade e potencial de diagnóstico. As evidências obtidas sugerem também que o grau de confiabilidade do *feedback* dos estudantes é superior ao de outras medidas (parte da explicação é que os estudantes acompanham os professores durante todo o ano e não somente durante visitas pontuais de avaliadores e que em geral significam notas dadas por 25-75 pessoas e não somente por um par de avaliadores).

Dentro desse segundo relatório encontramos também evidências sobre a confiabilidade dos avaliadores (controlando para nível educacional e experiência docente). Dentro de uma perspectiva na qual se espera que se leve de 17 a 25 horas para completar o treinamento dos avaliadores e taxas de acerto de 56% (para o MQI) a 83% (para o framework de Danielson), provou indispensável o papel do 'scoring leader' (um avaliador sênior). Claro, algumas especificidades desse estudo podem justificar esse cuidado, como o fato de que

os observadores desse estudo não tiveram nenhuma relação pessoal com os professores observados, dado que avaliaram vídeos de aulas. A mensagem final desse segundo relatório é que medidas combinadas de avaliação de professores fornecem melhores evidências do que medidas isoladas.

Dois relatórios do MET foram publicados em janeiro de 2013 na conclusão do projeto. O primeiro, assinado por Ho e Kane, aprofundou o tema do grau de confiança nas observações de classe feitas por profissionais da própria escola, examinando evidências resultantes da avaliação de 67 professores feita por 129 observadores que produziram umas 24 notas, totalizando 3.000 notas de vídeos. Várias questões fundamentais foram examinadas tais como, avaliações feitas por diretores da mesma escola vs diretores de outras escolas, os efeitos das avaliações feitas por outros professores, a influência da escolha dos professores sobre os vídeos a serem examinados, o número de vídeos alocados para cada examinador, o tempo de duração dos vídeos e o processo de alocação de vídeos aos avaliadores. Os principais resultados desse relatório foi que:

- Os observadores usam raramente as categorias extremas da avaliação (em um instrumento de 4 pontos);
- O poder de discriminação na avaliação dos diretores é 50% maior do que o dos outros professores;
- Mas os diretores avaliam seus próprios professores 0.1 mais do que diretores de outras escolas (também melhor que os professores em geral, que dão notas mais baixas, mas com menor variabilidade comparativa);
- Embora os rankings sejam similares aos rankings de outros diretores
- Quando os professores escolhem seus vídeos as notas são mais altas, mas de modo similar os rankings relativos são iguais a quando não puderam escolher os vídeos;
- As primeiras impressões dos avaliadores tendem a permanecer;
- Estima-se que a variação das avaliações de aula-a-aula para um dado professor (chamada em inglês de 'the lesson-within-teacher component') pode ir de 5 a 10% para diretores e 10-11% para outros professores. Diretores frequentemente avaliam os professores já com muita informação antecedente sobre eles;

- A melhor estratégia parece ser a de adicionar observadores sempre que possível (com melhores resultados mesmo quando o tempo de observação é reduzido para 15 minutos ao invés de uma aula inteira). Em outras palavras, adicionar avaliadores melhora a confiabilidade das medidas mais do que adicionar mais observações do mesmo avaliador. Assim, ter um vídeo visto por dois avaliadores diferentes produz resultados mais confiáveis do que ter um avaliador dando notas para dois vídeos do mesmo professor.

O segundo relatório final, assinado por Kane, McCaffrey, Miller e Staiger, também de janeiro de 2013, pergunta “Identificamos professores efetivos?” Nesse trabalho os autores colocam as medidas estudadas a teste com observações randomizadas para 1.181 professores, com outros 3.802 não randomizados e 17.153 professores não-participantes do projeto MET (mas cabe lembrar que os estudantes não foram randomizados em escolas diferentes). Alguns resultados importantes:

- Instrumentos de observação de classe e *surveys* de percepção dos estudantes têm menos poder preditivo do que medidas de valor-agregado;
- É importante controlar pelas notas de linha de base dos estudantes dado que esse componente não tem poder preditivo para resultados que seguem uma distribuição randomizada;
- Outra vez, medidas compostas ou mistas de efetividade dos professores tiveram melhor desempenho, também em outros testes

Esse último relatório realizou um grande esforço de organização e compilação de informação no processo de randomização. Parte da dificuldade, comum a esses processos de alocação randomizada, advém da necessidade de um plano de envio de calendários de aulas das escolas seguido da elaboração de um plano de alocação aleatória de avaliadores. Logo após foi necessário informar as escolas das alocações para elaboração e elaborar as bases de dados. No total o projeto solicitou informação sobre os horários de 2.462 professores em 316 escolas, criando 668 blocos de randomização.

Para concluir cabe mencionar que o projeto MET constitui uma fonte valiosa de informação sobre o impacto de diferentes instrumentos na avaliação de qualidade dos professores, que gerou e vem gerando muito debate e novas sugestões (Gargani e Strong, 2014). No entanto, tanto quanto o projeto oferece lições concretas importantes sobre aspectos logísticos e técnicos da aplicação desses instrumentos, ele é escravo de suas próprias premissas, como a de validar todas as medidas apenas com parâmetros de valor-adicionado. Com isso fica preso a uma perspectiva pedagógica específica, o que pode limitar o alcance de suas sugestões. Para que possamos entender isso melhor precisamos aprofundar a discussão sobre cada uma dessas medidas, incluindo a constituição de um universo mais amplo de instrumentos do que o oferecido pelo próprio ME.

## **2.2 Aprofundando os diferentes instrumentos de avaliação dos professores**

Para que possamos compreender os prós e contras de cada instrumento de avaliação devemos fazermos uma discussão mais detalhada dos mesmos. Em linhas mais específicas do que as anteriores dadas pelo agrupamento dos instrumentos em quatro grandes grupos, podemos classificar os principais, em 12 tipos e versões, nem todos mutuamente excludentes:

1. Modelos de Valor-Adicionado (“Value-Added”)
2. Modelos de Observação em Classe
3. Surveys de opinião de estudantes
4. Surveys de opinião de professores
5. Surveys de opinião de diretores
6. Modelo INFORM de observações de campo
7. Análise de *portfólio* dos professores
8. Modelos de simulação
9. Modelos de práticas reflexivas
10. Análise do conhecimento dos professores
11. Modelos de Auto eficácia
12. Modelos mistos

### **2.2.1. Modelos de Valor Adicionado**

Os modelos de valor-adicionado tentam medir o impacto dos professores através do desempenho dos alunos medidos por testes. A ideia aqui é simples: um bom professor é aquele que produz alunos e alunas que vão bem nas provas. Obviamente, o problema é saber se essas medidas de desempenho estão revelando o desempenho dos professores ou alguma outra característica não observável dos estudantes. Características não-observáveis são em geral intangíveis, como o QI de uma pessoa, ou sua capacidade de tolerância, ou alguma característica individual que não possa ser inferida ou observada pelas aparências. A maior parte dos estudos que discute esses estudos, como Bacher-Hicks et al (2017) é realizada dentro de um contexto não experimental (isto é, quando os dados são coletados em condições naturais). Isso é feito assim pois acredita-se que esses contextos quase-experimentais ou não experimentais contam com dados administrativos pré-existentes para diversos anos, o que possibilita estimativas mais precisas em comparação com as avaliações experimentais. Muito esforço também é dedicado para se controlar estatisticamente pelas características individuais dos estudantes (desempenho passado, background, etc) assim como pelas características médias de todos estudantes na classe (para levar em conta os chamados 'peer effects').

Para definir se as medidas de valor adicionado não são viesadas, normalmente se faz uma medição de base do desempenho dos alunos para dado professor seguido de uma medida no ano seguinte feita a partir de uma alocação randomizada de alunos para aquele professor (seguindo a metodologia de Chetty, Friedman e Rockoff, 2014).

É interessante, contudo observar que as medidas de valor adicionado (MVA) são normalmente atraentes para os gestores de política pública porque focam nos resultados e não nos insumos. Como tal elas são vistas como medidas objetivas. Além disso, como são construídas normalmente com dados não-experimentais, a partir de bases de dados já existentes, são quase sempre muito baratas em comparação com as alternativas. Por outro lado, as MVAs assumem que os professores têm influências isoladas no desempenho dos seus alunos (ou seja, a matemática aprendida por um aluno vem somente do seu professor de matemática), o que pode não acontecer. Similarmente, testes com notas máximas atingíveis de realização (isto é, que sejam fáceis de serem atingidas) podem não discriminar



entre os melhores professores. Também deve ser mencionado que as MVAs se aplicam mais facilmente a poucas matérias mais sujeitas à quantificação, como matemática ou português (“literacy” do inglês). Talvez a crítica mais contundente seja a de que as MVAs medem apenas os resultados das crianças nos testes, mas que isso não necessariamente vai ser equivalente ao seu aprendizado nem tampouco reflita uma avaliação precisa do trabalho do professor. Adicionalmente, professores podem ‘ensinar para o teste’ assim como ‘alunos e diretores de escola’ podem trapacear para ver sua escola subir no ranking. Do mesmo modo, o que muitas vezes se mede em provas não é o resultado do trabalho do professor, mas o resultado das interações entre o professor e seus alunos, como alerta Papay (2012). Por fim, como argumenta Ballou (2009), medidas de valor-agregado são vulneráveis a resultados que não são produzidos em escalas com intervalos constantes e o uso da teoria da resposta ao item (TRI) não parece resolver esse problema.

Dentro da literatura que discute aspectos econométricos dos MVAs, tal como Araujo et al (2016) e Ishii e Rivkin (2009), vemos que podemos dizer muito pouco com somente um ano de dados. Por exemplo, não podemos separar as diferenças de qualidade dos professores de choques aleatórios no nível das classes (como por exemplo, a presença de um estudante particularmente difícil que atrapalha sistematicamente a classe, ou o relacionamento de classe daquelas crianças naquele ano, ou eventos como a violência ou ondas de gripe). MAVs capturam muito pouco as qualidades multidimensionais e as complexidades de ser professor. Pouco ainda sabemos sobre sua confiabilidade e validade e pouco tem ajudado para o estabelecimento de lições robustas sobre como podem ajudar na identificação e formação de professores de qualidade (Hanushek e Rivkin, 2010). Também devemos levar em conta aspectos técnicos como o chamado ‘ceiling effect’ (ou o efeito das escalas de nossas medidas que podem ter tetos diferentes) como nos mostra Koedel e Betts (2010).

### **2.2.2 Modelos de Observação em classe**

Os modelos de avaliação de classe são mais específicos que os modelos de valor adicionado pois começam com uma visão pedagógica que se materializa em instrumentos pedagógicos e currículos bem definidos que são testados através de observações da

prática em sala-de-aula. Algumas vezes o mais difícil é a tradução de elementos de observação, de natureza etnográfica em um conjunto de procedimentos simples que possam ser aplicados por professores (Baecher e Kung, 2011). Uma lista dos principais instrumentos inclui:

**a)** A Abordagem para o Ensino (“Framework for Teaching” ou FFT), desenvolvida por Charlotte Danielson: na sua última edição de 2013, esse instrumento, categorizado em quatro dimensões (Planejamento e preparação; o Ambiente de Classe; Instrução e Responsabilidades profissionais) fornece um conjunto de rubricas detalhadas dentro das quais é possível classificar diferentes níveis de proficiência (insatisfatório, básico, proficiente, com distinção). A Tabela 1 abaixo ilustra uma sequência de rubricas para o domínio 1 de demonstração do conhecimento de conteúdo e pedagogia. Independente de concordarmos ou não com essas rubricas particulares fica claro que o trabalho de avaliação de professores dentro desse modelo começa com a elaboração de rubricas onde são registradas as expectativas dentro de cada critério de qualidade. Há uma hipótese implícita de que os elementos das rubricas estão relacionados entre si, mas isso poderia não acontecer adicionando complexidade à aplicação das mesmas (ou seja, imagine um professor que erre conteúdos e que, portanto, seria do nível insuficiente consiga citar relações de conteúdos interdisciplinares –uma característica de um nível de distinção.)

**Tabela 1 – Exemplos de atributos críticos nas rubricas de FFT/Danielson**

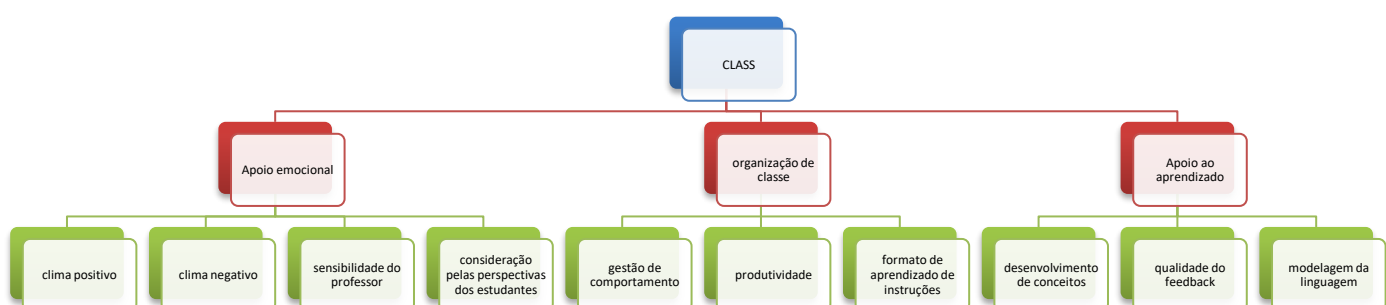
<b>Níveis</b>	<b>Atributos críticos</b>
<b>Insuficiente</b>	<ul style="list-style-type: none"> <li>- Erros de conteúdo</li> <li>- Desconsideração de relações de pré-requisitos no planejamento de classe</li> <li>- Uso de estratégias inapropriadas para a disciplina</li> </ul>
<b>Básico</b>	<ul style="list-style-type: none"> <li>- Entendimento da disciplina é rudimentar</li> <li>- Conhecimento de relações de pré-requisitos é impreciso ou incompleto</li> <li>- As lições usam estratégias limitadas de instruções e algumas não são adequadas aos conteúdos</li> </ul>

<b>Proficiente</b>	<ul style="list-style-type: none"> <li>- Identificação de importantes conceitos da disciplina e suas relações</li> <li>- Explicações claras do conteúdo</li> <li>- Respostas precisas as respostas dos estudantes e feedback que melhora seu aprendizado</li> <li>- Os planos de aula e estratégias de instrução são completamente adequadas aos conteúdos</li> </ul>
<b>Distinção</b>	<ul style="list-style-type: none"> <li>- Citação de relações de conteúdo intra e interdisciplinares</li> <li>- Planos demonstram consciência de possíveis erros dos estudantes e de como eles podem ser resolvidos</li> <li>- Planos refletem desenvolvimentos recentes em pedagogia relacionada aos conceitos</li> </ul>

Fonte: Danielson (2013)

**b)** O Sistema de Scores de Avaliação em classe (“Classroom Assessment Scoring System” ou CLASS), desenvolvido por Robert Pianta, Karen La Paro e Bridget Hamre da Universidade de Virgínia, é sem dúvidas o mais conhecido desses sistemas de observação de classe. A sua estrutura pode ser descrita através do diagrama 1 abaixo:

Diagrama 2 – Estrutura analítica do CLASS



A dimensão de apoio emocional mede a conexão emocional entre os professores e estudantes e a conscientização dos professores e respostas aos problemas dos

estudantes. A organização de classe mede o grau no qual os professores lidam de modo proativo com problemas comportamentais dos estudantes e dão a eles uma oportunidade de aprendizado. Por fim, o apoio ao aprendizado mede o grau no qual os professores promovem os entendimentos conceituais dos alunos através de um feedback orientado pelos processos e com o uso de linguagem apropriada a sua idade.

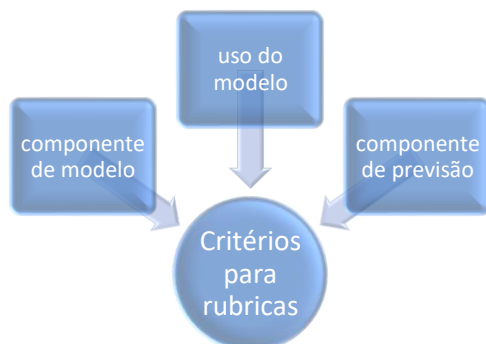
Conforme descrevem Curby et al (2012) para operacionalizar o CLASS é necessário um treinamento dos pesquisadores de pelo menos 2 dias para que esses se tornem codificadores (“coders”) confiáveis do CLASS. O objeto do treinamento é na revisão do manual do CLASS, incluindo mostras de vídeos de classes e discussões sobre como classificar potencialmente aquelas classes. No final do treinamento é dado um teste e os participantes tem que codificar vídeos de 20 minutos. Os codificadores classificados como ‘confiáveis’ são aqueles que conseguem codificar mais de 80% das situações. No estudo original de Curby et al (2012), conduzido em Washington nos Estados Unidos, pares de avaliadores visitaram cada classe, mas sem avisar. Ou seja, os professores não sabiam que dia específico os pesquisadores estavam vindo. No entanto, ao chegar os pesquisadores sempre perguntavam aos professores se aquele dia era um dia típico. Se alguma atividade especial tivesse sido planejada para aquele dia, os pesquisadores voltariam outro dia. Cada observação de 2 horas consistiu de 4 ciclos de observação, com cada ciclo consistindo de uma observação de 20 minutos seguida de 10 minutos de codificação.

As medidas de observação CLASS (cujas origens estão no NICHD – “Study of Early Child Care and Youth Development”) são feitas através da geração de diversos vídeos e áudios. Baseados em rubricas (o segredo parece ser a geração destas), CLASS trabalha com codificações que fornecem um guia concreto de avaliação nos quais os professores são classificados em escores ‘baixo’ (notas 1-2), ‘médio’ (notas 3-5) e ‘alto’ (notas 6-7). Na prática os escores são altamente correlacionados entre si e por isso muitas vezes se trabalha com um escore CLASS total. Protocolos como CLASS podem dar um retorno importante aos professores. De fato, podem ajudar inicialmente com um conjunto de padrões de instrução concreto e com rubricas detalhadas que explicam aos professores

como podem ser avaliados. Dentro do DNA dessa classe de medidas há uma preocupação grande com o gerenciamento das interações com os estudantes e em cenários de respostas dos estudantes ao professor. Ao traduzir o ambiente e a prática da sala-de-aula em indicadores, CLASS dá evidências mais facilmente entendidas e mais concretas aos professores. Mas esse não é um ponto consensual na literatura. Para que o CLASS possa ser aplicado sua confiabilidade precisa ser garantida o que demanda principalmente o treinamento e a certificação dos avaliadores (do inglês, “raters”). O que parece, no entanto, ficar claro é que avaliações do tipo CLASS podem ser vistas como instrumentos de desenvolvimento profissional e essencial a ‘avaliação formativa’ dos professores.

É interessante notar também modelos de observação de classe baseados no uso de vídeos tais como o de Norton, McCloskey e Hudson (2011). Nessa proposta se usa um instrumento chamado ‘construção de modelos’ que se refere as atividades correntes e interativas dos professores de provisão de explicações razoáveis para as ações dos estudantes (incluindo suas verbalizações). No caso da matemática, esses modelos simulam as ações mentais que os estudantes desempenham quando resolvem problemas matemáticos e o poder de um modelo é determinado pelo seu potencial explicativo e preditivo. Há então uma discussão sobre os ‘modelos cognitivos das estratégias de resposta dos estudantes’. Nesse modelo de observação de classe, são mostrados vídeos aos professores de seus próprios estudantes resolvendo problemas com a pergunta de que eles prevejam as estratégias de resposta dos estudantes e seus acertos ou erros. Uma vez feito isso os pesquisadores medem a correlação entre os acertos dos professores em prever as respostas dos alunos e seus acertos em testes que medem suas habilidades. Esse projeto levou ao desenvolvimento de rubricas em três categorias:

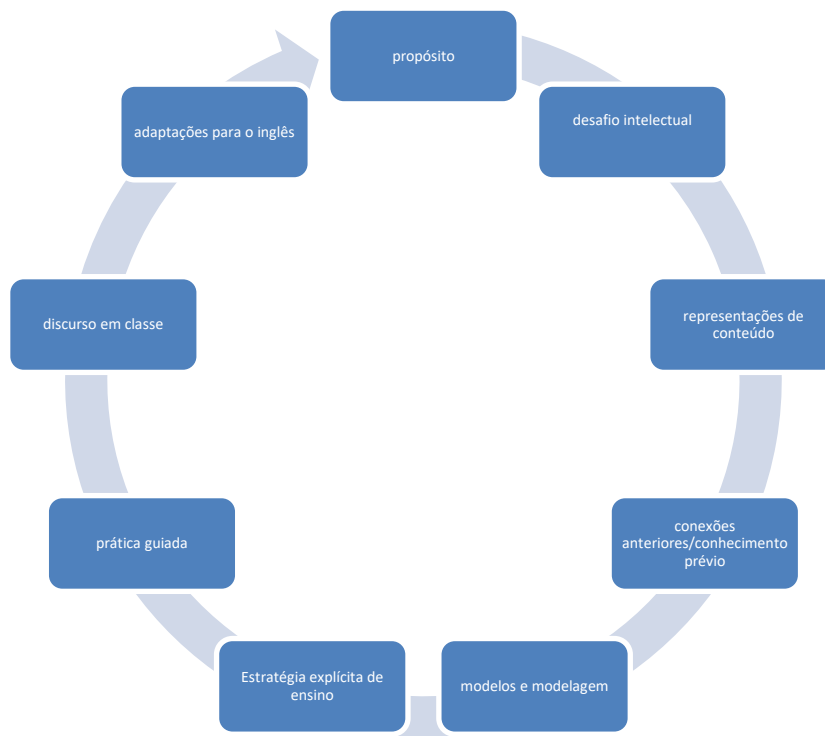
### **Diagrama 3 – Critérios para rubricas**



O ponto principal da análise consiste no estabelecimento da confiabilidade e validade desses instrumentos. Para isso é usado muitas vezes o chamado *Kappa de Cohen*, que é uma medida de acordo entre avaliadores para itens qualitativos (categóricos). Essa medida é mais robusta do que um simples cálculo de acordo percentual. Os avaliadores (“raters”) classificam  $N$  itens em  $C$  categorias mutuamente exclusivas (Smeeton, 1985).

**c)** O Protocolo para Observações do Ensino de Linguagem (“Protocol for Language Arts Teaching Observations” ou PLATO), desenvolvido por Pam Grossman da Universidade de Stanford, é outro dos *instrumentos de desenvolvimento profissional*, como são assim chamados, entre os mais conhecidos. PLATO foi concebido para ser aplicado ao longo de uma variedade de currículos e abordagens pedagógicas. Grossman, Greenberg, Hammerness, Cohen, Alston e Brown (2009) utilizam o protocolo para gerar evidência sobre o impacto de algumas práticas em sala-de-aula (como por exemplo a qualidade do discurso usado pelo professor) no aprendizado dos alunos e desenvolvimento de seu entendimento conceitual e habilidades de raciocínio (nem sempre captadas por notas de testes padronizados). Desse modo, PLATO foca mais em componentes de aprendizado ou instrução do que em aspectos de gestão escolar. Em outras palavras, PLATO olha detalhadamente para os elementos da prática de sala-de-aula relacionados ao desempenho dos alunos (diagrama 3) e mostra como a inclusão de medidas de observação de classes são necessárias para medir a efetividade e qualidade dos professores.

**Diagrama 4 – Elementos do PLATO**



É interessante notar como a medida de cobertura de conteúdo não parece diferenciar entre professores, mas sim tempo gasto em pesquisa. Também parece ser que professores de grupos de baixa e alta qualidade agrupam seus estudantes de modo diferente: os melhores professores trabalham com grupos menores de alunos (Grossman, Loeb, Cohen e Wyckoff , 2013) .

**d)** A Qualidade Matemática da Instrução (“Mathematical Quality of Instruction” ou MQI), desenvolvido por Heather Hill da Universidade de Harvard. O ponto de partida desse instrumento de avaliação é a natureza do conteúdo matemático disponível para os estudantes, visto através de 3 ângulos, quais sejam: relação professor-aluno, interação entre professores e conteúdos e interação entre alunos e conteúdos. Um aspecto interessante do MQI, que pode ser generalizado, é que ele foi criado com o entendimento de que o conhecimento matemático possui algumas especificidades que não podem ser adequadamente capturadas por instrumentos gerais, como o CLASS.

Assim, atenção ao conhecimento pedagógico ou ao estilo dos professores ou mesmo o clima da classe, apesar de importantes, não são suficientes para capturar aspectos que são importantes para o ensino da matemática na sala-de-aula como o feito pelas dimensões acima (Hill et al, 2008). Essa medida é desse modo estruturada em cinco dimensões tradicionalmente: i) erros e imprecisões, ii) riqueza do ensino da matemática, iii) trabalho com os estudantes, iv) relação do trabalho de classe com a matemática e v) práticas com os estudantes alinhadas com o currículo básico. No entanto, outras dimensões aparecem como vi) participação dos estudantes na produção dos significados matemáticos e vii) completude e grau de clareza em algumas outras publicações usando MQI (mais informações podem ser vistas diretamente no site de Harvard, <https://cepr.harvard.edu/mqi-domains>). As avaliações são feitas em segmentos de igual tamanho de 5 ou 7.5 minutos. Em geral são utilizados dois avaliadores que classificam a classe dentro dos cinco domínios principais do MQI (existem problemas de confiabilidade dos avaliadores que devem sempre ser levados em conta segundo Hill et al, 2012).

O MQI constitui uma rubrica importante em si mesma, mas que sinaliza a importância de atentarmos para especificidades pedagógicas relacionadas a conteúdos acadêmicos específicos, com instrumentos correspondentes. Dentro de uma perspectiva de modelos mistos e em função dos resultados conseguidos pelo MET essa é uma lição que não deve ser desconsiderada. Vale o mesmo para instrumentos como o QST (“Qualidade do Ensino da Ciência”) ou outros formulados para avaliar o conhecimento tecnológico dos estudantes (Baniester e Reinhart, 2012) que não serão discutidos aqui pois apenas traduziriam a mesma mensagem que o MQI.



e) O Protocolo de Observações de Professores (“UTeach Teacher Observation Protocol” ou UTOP), desenvolvido por Michael Marder e Candace Walkington da Universidade de Texas-Austin em 2009 foca sua atenção no ensino de ciências e matemáticas (áreas onde já falta professores nos Estados Unidos) com o mesmo argumento estabelecido pelo MQI de que precisamos instrumentos específicos para áreas específicas do conhecimento. Surge como uma crítica a instrumentos prévios de observação que não reconheciam a importância de conteúdo específicos dessas áreas. Está organizado em quatro grandes dimensões (ambiente de classe, estrutura da lição, implementação e conteúdos de ciência e matemática) que incluem 32 indicadores. O UTOP também inclui uma entrevista com os professores, feitas após as observações, que contém 12 questões abertas, para confirmar o *background* e os propósitos das classes. O UTOP completo contém também formulários demográficos para professores e avaliadores.

Os avaliadores são treinados a partir de um ‘Manual de Treinamento UTOP’ que detalha os critérios de classificação para cada item. As observações são feitas independentemente em um primeiro momento entre pares de avaliadores, seguidas de um ‘debriefing’ que é um elemento importante nesse instrumento de avaliação. A ideia é de que pelo menos um dos avaliadores seja um especialista em uma das áreas avaliadas, ou seja, matemática ou ciências. Cada observação pode durar 50 ou 90 minutos, seguida de entrevistas pós-observação que podem ser feitas presencialmente, por telefone ou por e-mail. Dentro do protocolo sugerido as entrevistas devem ser gravadas e transcritas.

Do ponto de vista analítico é feita uma categorização das entrevistas de acordo com o background das escolas e professores, segundo as matérias ensinadas, tipos de escola, nível de pobreza das escolas, nível das classes e de experiência dos professores. Os observadores usam uma escala Likert de 7 pontos para classificar 32 indicadores. A consistência interna das medidas é verificada através do alpha de Cronbach. O estudo de Walkington et al (2011) mostra como os fatores de background (chamados ‘pre-service characteristics’ em inglês) são determinantes na formação e desempenho de professores de ciências e matemática.

Esses são os instrumentos de observação de classe mais discutidos na literatura e que são considerados importantes por aportarem sugestões aos professores de como podem melhorar suas classes (diferentemente de medidas de valor-agregado que pouco elucidam as causas no âmbito do trabalho dos professores com bom ou mal desempenho de seus alunos). Mesmo sendo construídos a partir de fundamentações e ênfases pedagógicas distintas têm em comum a elaboração de protocolos de treinamento, de rubricas, de critérios de avaliação, de homogeneização de processos de interação e de reflexão que são muito úteis para a criação de novos instrumentos de observação.

### **2.2.3. Surveys de opinião de estudantes**

Sobre os surveys de estudantes, o mais utilizado e reconhecido na literatura é o chamado 'Tripod Survey', desenvolvido por Ron Ferguson da Universidade de Harvard, que é uma lista de 26 itens baseados na escala Likert que cobre questões relativas a como os estudantes se sentem, por exemplo, em relação ao nível de engajamento em classe, ou clareza da aula do professor, ou se o mesmo dá uma explicação útil, ou se sentem se a classe é produtiva, entre outros pontos. O Tripod já está na sua 18ª edição, utilizando uma estrutura chamada os '7 Cs do Ensino Efetivo', incluindo, cuidar, esclarecer, cativar, consolidar, reunir, desafiar e gerir as classes (do inglês, "care, clarify, captivate, consolidate, confer, challenge e classroom management"). Atualmente quase todo o processo é feito de modo comercial e online (<https://tripoded.com>). Mais especificamente as dimensões tratam de:

- Cuidar: a melhor tradução seria 'se importar'. Aqueles professores que se importam com os seus alunos lhes dão apoio emocional e se interessam pelo o que eles fazem;
- Reunir: professores que se reúnem falam com os estudantes assim como acolhem e respeitam suas perspectivas
- Cativar: professores que cativam os seus alunos fazem o aprendizado interessante e relevante
- Esclarecer: explicar os assuntos com clareza, oferecer feedback, diminuir dúvidas para que os alunos possam entender as lições
- Consolidar: resumir e integrar o aprendizado dos alunos

- Desafiar: consiste em pressionar os estudantes para que possam pensar sobre os conteúdos de forma rigorosa, com resiliência
- Gerir a classe: promovendo um clima de classe cooperativo e respeitoso

Diversos desses componentes, como esclarecer e desafiar, têm sub-dimensões, de tal modo que esses indicadores representam índices compostos. Cabe, no entanto, notar que a base informacional do TRIPOD é constituída pela visão dos alunos que pode estar sujeita a uma ampla variedade de fatores, para não mencionar que muitas vezes os alunos podem não estar aptos para avaliar a dimensão pedagógica do trabalho de um professor. Outro problema, discutido por Backer (2012) é que os estudantes podem penalizar os professores em retaliação a notas baixas. Seja como for, parece que avaliações como o TRIPOD fornecem uma relevante fonte de informação complementar a outros instrumentos, demandando muita cautela interpretativa no caso em que seja utilizado como único instrumento.

Evidências empíricas obtidas por Champagne (2013) sugerem que embora a taxa de resposta dos estudantes entre diferentes meios (questionários em papel, online em computadores, online em telefones, etc) possa ser diferente, não há diferença estatística relevante entre os rankings de notas dadas aos professores.

#### **2.2.4. Surveys de opinião de professores**

Existem vários modelos de questionários que procuram captar a opinião dos professores sobre o significado de efetividade no seu trabalho assim como das condições necessárias para um bom ambiente de trabalho. Uma das referências mais importantes na literatura é dada por Stronge (2007) que categorizou as principais qualidades:

**Tabela 2 – Qualidades dos Professores Efetivos**

Categoria	Qualidade
Características pessoais	Preocupação com os estudantes Tratamento dos estudantes com justiça e respeito Entusiasmo ao ensinar Motivação para trabalhar Devoção ao desenvolvimento profissional Conscientização dos lados mais fortes e mais fracos
Gestão de classe	Estabelecimento de gestão de classe Organização de tarefas, instrumentos e espaços em classe Estabelecimento de disciplina entre os estudantes
Planejamento dos materiais	Consideração que a preparação dos materiais é importante Gerenciamento de tempo Possuir altas expectativas para si mesmo/a e seus estudantes Organização de conteúdos para uma instrução efetiva
Instrução	Uso de diferentes métodos e estratégias de instrução Orientação dos estudantes com exemplos claros Promoção do aprendizado por entendimento ao invés de memorização

	Uso do questionamento em sala-de-aula de modo efetivo Direção da atenção do estudante para a lição
Monitoramento do progresso dos estudantes	Dar o tema apropriado ao conteúdo da lição e à capacidade do estudante Propiciar interação face-à-face aos estudantes que ficam para trás na classe Considerar as necessidades dos estudantes e sua proficiência

É importante mencionar que não devemos esperar que todos os professores tenham a mesma opinião. Por exemplo, em uma aplicação desse modelo para a Turquia, Bozkus e Tastan (2016) mostraram que há uma divergência significativa de visão entre os professores já estabelecidos e aqueles que começam na profissão. Para aqueles que iniciam na profissão, a característica mais relevante para definir a qualidade de um professor é o seu *conhecimento*. No entanto, para os professores já estabelecidos é a *gestão de classe* o elemento determinante de sua qualidade com o maior impacto sobre o progresso dos estudantes. Esses efeitos podem ser diferentes também quando tratamos de avaliar classes nas quais temos professores e professores assistentes. Curby et al (2012) mostraram que os professores assistentes, em comparação com os professores líderes, têm interações de pior qualidade com os estudantes no domínio do apoio do aprendizado, mas que ambos conseguem trabalhar algumas vezes de modos sofisticados com dinâmicas difíceis de serem captadas. O que parece fazer diferença é a experiência dos professores. De qualquer modo a evidência sugere que professores assistentes e líderes não são diferentes em relação ao tipo de apoio emocional e organização da classe. Mas essa permanece uma importante questão empírica a ser verificada dependendo do contexto.

Evidências produzidas no Reino Unido sugerem que uma cultura de avaliação de professores pode ajudar no aperfeiçoamento das habilidades de auto-avaliação dos professores que estão se preparando ou iniciando a carreira (Gossman e Horder, 2016).

### 2.2.5. Surveys de opinião dos diretores

Modelos de avaliação baseados na visão de diretores, tais como sugerido por Jacob e Lefgren (2008), mostram que diretores podem em geral identificar professores que produzem as piores e melhores resultados (*top* e *bottom* 10%-20%), mas que são menos capazes em distinguir entre aqueles professores com níveis intermediários de eficiência. Em todo o mundo há uma resistência dos professores em serem avaliados, fundamentada em uma preocupação com a falta de capacidade dos diretores (ou outros membros da administração) em aplicarem uma avaliação justa (Murnane e Cohen, 1986).

São vários os fatores que podem explicar alguns vieses na visão dos diretores, tais como sua idade, seu gênero ou sua proximidade (do inglês, "likability") dos professores avaliados. O potencial de distorção de informações puramente subjetivas é conhecido na literatura das capacitações, como discutido exaustivamente por pensadores como Amartya Sen e Martha Nussbam.

Normalmente, um ordenamento de atributos de professores feitos por diretores poderia incluir os seguintes critérios:

- Efetividade geral do professor
- Dedicção e ética do trabalho
- Organização
- Gestão de classe
- Melhoria das notas dos estudantes nas matérias X, Y, Z
- Modelo pessoal para os estudantes (ser um exemplo)
- Satisfação dos estudantes com o professor (como vista pela direção)
- Satisfação dos pais com o professor (como vista pela direção)
- Relacionamento positivo com outros professores (colegas)
- Relacionamento positivo com administradores

O protocolo básico é operacionalizado através de ordenamentos, feitos pelos diretores, que são categóricos e que não possuem uma interpretação cardinal. Existe toda uma literatura de medidas não-paramétricas de associação entre indicadores de desempenho feitas

através de processos de simulação de Monte Carlo (métodos estatísticos derivados de amostragens aleatórias com grandes números). Como ponto final cabe mencionar que apesar de todos seus méritos temos visto uma tendência contrária à utilização desse método, que possivelmente está na contramão de métodos mais contemporâneos baseados na auto-reflexão dos professores.

### **2.2.6. Modelo INFORM de observação de campo**

O modelo INFORM (do inglês, “Instructional Field Observations Rounds Model), introduzido por Hoyt e Terantino (2015) é baseado no conceito de rounds que residentes médicos têm em hospitais com médicos experientes. Essa ideia foi originalmente proposta por City et al (2009), mas o modelo INFORM foca em estudantes de pedagogia que se preparam para dar classes. O modelo tem como base visitas de grupo a estudantes do fundamental seguido por sessões chamadas de “debriefings” que incluem os estudantes de pedagogia, seus professores e os professores das escolas que visitaram.

Não é que não existam modelos de visitação de escolas em cursos de pedagogia. Mas é que frequentemente os estudantes de pedagogia são convidados a adotar uma atitude inteiramente passiva (muitas vezes evitando qualquer interação com os professores observados) e tradicional nessas atividades. Isso não permite um esforço ativo por parte dos alunos de conexão entre as teorias que aprendem e as práticas que vivenciam.

O modelo INFORM busca assim ser um instrumento de coleta de observações de campo baseado em um exercício de colaboração entre professores de ensino fundamental, estudantes de pedagogia (no inglês, a referência é feita a “pre-service teacher candidates”) e professores universitários. Podemos representar esquematicamente suas etapas do seguinte modo:

### **Diagrama 5 – etapas do INFORM**



Do ponto de vista operacional, inicia-se com a identificação e recrutamento de escolas e professores colaboradores. Depois, deve-se pensar o portfólio de escolas, diverso o suficiente, para representar o contexto proposto. Questões referentes à disponibilidade de turmas e horários e deslocamento de alunos devem também ser enfrentadas. Essencial, contudo, é em termos um roteiro de práticas essenciais que ajudem os alunos no seu processo reflexivo. Os alunos devem ser preparados para discernirem entre linguagem descritiva vs linguagem normativa, com um foco em práticas pedagógicas e sensibilidade para diferenciar entre questões diferentes, como o conteúdo de uma lição vs interação de estudantes.

O que chamamos de 'debriefing' é a alma do processo pois é o componente que dá aos candidatos/as a professor/a um acesso direto ao entendimento do funcionamento dos métodos pedagógicos na prática e todas as sequencias distintas de processos de tomada de decisão que são observadas durante as visitas. O método propõe que através dessas conexões os alunos conseguem detectar a intencionalidade do planejamento pedagógico fazendo assim com que tenha um modo de implementação mais efetivo. O texto de Hoyt e Terantino (2015) contém os dois guias para os estudantes de pedagogia/professores, o primeiro um guia de observação de aulas e o segundo um guia para o relato reflexivo.

### **2.2.7. Análise de portfólio dos professores**

O portfólio é o nome dado ao conjunto de documentos e outros artefatos que os professores têm para a preparação de suas classes. Aqui podemos incluir programas de curso, guias, exemplos de materiais distribuídos ou de testes feitos, notas de aula, apresentações,



resumos de avaliações de estudantes, relatos feitos por outros professores que já observaram suas classes ou mesmo exemplos de trabalhos feitos pelos seus alunos.

Mas um portfólio é algo mais do que isso porque além de ser uma coleção de materiais inclui uma narrativa dos professores explicando como elaborou e como utilizou cada um desses materiais. Isso de algum modo pode ser visto como um momento de auto-reflexão dos professores (Chism e Stanley, 1999). Esse portfolio pode então ser avaliado por colegas ou administradores. Portfolios podem ser considerados um tipo de informação mais objetiva em relação a surveys de opinião pois dão materialidade ao trabalho dos professores durante um fluxo definido de tempo.

No entanto, pesquisa recente feita por Henderson, Turpen, Dancy e Chapman (2014) mostrou que normalmente portfolios de professores não são considerados um método de avaliação da efetividade dos professores. Muito menos são considerados (assim como avaliações provenientes de surveys de estudantes) como elementos para a promoção dentro da carreira. Se por um lado esse método pode fornecer informações objetivas, por outro ainda está muito incipiente e não possui um protocolo ou rubricas específicas que podem ser usadas na sua administração.

### **2.2.8. Modelos de Simulação**

Sistemas de avaliação de professores baseados em simulações aparecem em duas modalidades principais, i) simulações de vídeos e ii) simulações com software. Ambas fornecem aos professores experiências e oportunidades de discussão de situações de classe, incluindo questões pedagógicas ou de gestão de sala-de-aula. O primeiro aspecto importante em simulações de vídeo, que em geral são mais complexas do que simulações por computadores, é a identificação do que está caracterizando determinada situação em classe. Depois, passa-se ao estabelecimento de conexões entre eventos específicos e princípios mais gerais de ensino e aprendizado. Treinamentos por simulação possibilitam a

discussão de decisões caracterizadas por um exato momento dentro da sala-de-aula. Parece fácil, mas nem sempre a identificação do que importa ou não é algo trivial, como nos explica Barbara Herman no seu livro *Moral Literacy*.

Modelos de simulação estão alinhados com uma preocupação com o 'ensino da prática' (Grossman et al, 2009) que incluem não somente um processo de representação dessas práticas, mas de decomposição e aproximações ao mundo real. Do ponto de vista da representação, a elaboração dos vídeos propicia uma oportunidade de engajar educadores na experiência de co-criação de roteiros e de desenhos pedagógicos de implementação prática de estímulo a certos processos de pensamento dos alunos.

Existem de fato muitas atividades baseadas nessa preocupação com o ensino da prática que procuram detalhar as ações pedagógicas dos professores durante as classes. Uma dessas práticas comuns é chamada de *microteaching* (a tradução do inglês seria algo como 'micro-lições') na qual um professor dá uma aula ou uma parte de uma aula para um pequeno grupo de estudantes (ou de outros professores) em um contexto com tempo e elementos reduzidos assim como com resultados de aprendizado simplificados.

Simulações de vídeos podem ser úteis para a prática de aulas (do inglês, 'rehearsing teaching') nas quais os professores podem avaliar a estrutura de suas aulas a partir de modelos pedagógicos ou técnicas pedagógicas distintas (Lampert et al, 2013).

O mais interessante no modelo de simulações com vídeo, contudo, não está do lado dos professores, mas sim dos alunos. É precisamente na leitura de sua classe (e no acompanhamento do aprendizado e processo de raciocínio de seus alunos) que um professor pode mostrar sua efetividade como um educador. Em outras palavras, entender como os estudantes pensam e como erram é um atributo essencial na bagagem de um bom professor (Jacobs et al, 2010).

Um dos aspectos mais inovadores dessa discussão sobre simulações é a análise de padrões de resposta (do inglês, “response patterns”) entre professores e alunos. A codificação de roteiros, nos quais por exemplo, um professor pode dar ao aluno o contexto adequado de aprendizado que pode leva-lo ao reconhecimento do seu erro e da resposta correta, é um dos grandes desafios dessa modalidade de avaliação. Amador (2017) mostra como critérios tais como i) foco, ii) expressões, iii) temporalidade e iv) padrões discursivos podem ser usados para a análise de diferentes simulações. Por exemplo, professores que colocam questões abertas a seus estudantes estimulam seus processos de pensamento em contraposição a outros professores que também fazem perguntas, mas que são apenas do tipo ‘sim’ ou ‘não’.

Na simulação por computadores grande ênfase é dada a mensuração de habilidades de ordem mais elevada (do inglês, as chamadas “higher-order skills) cujo propósito é fomentar a interatividade, adaptação e utilidade nos processos educacionais. Um dos modelos mais utilizados é o *SIM school* (<http://www.simschool.org>) que é um simulador de um ambiente de classe que recria situações desenhadas para a tomada de decisão dos professores (junto com representações de modelos de aprendizado). O modelo contém domínios cognitivos e psicológicos que seguem a *Taxonomia de Objetivos Educacionais de Bloom* (Bloom, Mesia e Krathwohl, 1964) combinados com modelos de psicologia de McCrae e Costa (1996).

O simulador atribui valores entre -1 e +1 as propensões de processamento emocional dos professores, que podem mudar de acordo com a mudança de domínio e contexto educacional. De modo simples pode-se dizer que o programa coleta dados automaticamente à medida que os professores interagem com os estudantes virtuais e tomam decisões sobre o que e quando ensinar. Uma das vantagens operacionais desse ambiente virtual é o uso em plataformas de telefones celulares (Gibson, 2013).

### **2.2.9. Modelos de Práticas Reflexivas**

Modelos de práticas reflexivas podem ser considerados uma extensão dos modelos de observação de classes que procuram resolver o problema de falta de internalização e apropriação de novos modelos pedagógicos. Como tal procuram adicionar valor não somente ao dar um feedback aos professores (como fazem modelos do tipo CLASS) mas

ao preocuparem-se com o processo de reflexão e internalização de novas práticas. Dito de outra forma, para que os professores possam adotar efetivamente novas práticas pedagógicas é preciso que eles tenham uma reflexão ativa crítica de seu próprio trabalho (de modo isolado ou conjunto com outros professores) (Hawley e Valli, 1999; Porter et al, 2003).


Além disso é importante que essas experiências sejam desenvolvidas dentro dos contextos escolares ao invés de outros contextos que servem de ambiente para o treinamento de professores, como hotéis ou prédios públicos administrativos. A transferência e aplicação de conhecimento depende muito do contexto no qual ele é adquirido e aplicado. Essa não é uma proposta nova, podendo ser encontrada nos escritos de Dewey (1933). Mas no contexto em que ela é sugerida pode representar uma demanda importante para as avaliações e programas de treinamento de professores, no sentido específico de que as avaliações devem estimar se os professores possuem ou não uma capacidade reflexiva de suas próprias práticas como condição essencial para sua mudança de atitude. Um tipo de contexto relevante para a realidade brasileira é o dado por estudantes com problemas disciplinares. Evidências com experimentos na Holanda mostraram que adaptações de instrumentos para contextos com problemas comportamentais não é uma tarefa trivial (Buttner et al, 2015).

Uma complicação nessa perspectiva é que a capacidade reflexiva dos professores pode depender não de características individuais, mas de seu pertencimento a redes sociais específicas. Nesse contexto, o conhecimento teria uma dimensão pública difícil de ser combatida individualmente. Assim, os principais passos seriam:

## **Diagrama 6 – Modelo reflexivo**

 internalização  
Apropriação

 personalização  
Transformação

 exame crítico  
de crenças,  
ações e  
propósitos  
Reflexão

Se os professores não se apropriarem e fizerem sentido dos desafios e dilemas propostos pelas avaliações eles possivelmente serão um obstáculo e não um catalizador das mudanças desejadas. Evidências empíricas deste modelo, como Camburn (2010) sugerem através do uso de modelos multi-nível que práticas reflexivas são influenciadas por interações sociais que ocorrem em oportunidades de aprendizado na escola em comparação com aquelas resultantes de formatos mais tradicionais. As evidências sugerem também que há uma menor incidência de práticas reflexivas em contextos de escolas com alto nível de privação. Não deve causar surpresa que professores trabalhando em condições difíceis, de isolamento muitas vezes, sem tempo para reflexão, com classes disfuncionais e em ambientes de violência enfrentam barreiras quase intransponíveis. Uma outra lição importante é que o aprendizado dos professores ocorre através de uma série de interações ao longo do tempo, comprovando a baixa efetividade de treinamentos do tipo “one-off”.

### **2.2.10. Análise do Conhecimento dos professores**

Apesar de relativamente óbvia, mas pouco utilizada, uma maneira direta de avaliar a qualidade dos professores e de seu ensino, é testar diretamente o conhecimento dos professores. Não é necessário limitar esses testes a conhecimentos puramente teóricos, pois práticas pedagógicas e de gestão de classe também podem ser testadas, como por exemplo através de instrumentos como o Survey de Professores de Matemática do Ensino

Fundamental da Universidade de Michigan (no inglês, “Multiple-Choice Elementary Teacher Survey”). Charalambous (2016) testou esse modelo para Chipre focando em 4 eixos:

- Provisão e avaliação de explicações: trata-se aqui de registrar como os professores respondem aos alunos, se promovem um ambiente de classe participativo e o uso da razão e se os próprios alunos oferecem explicações quando perguntados
- Seleção e uso de representações: avalia-se o uso e grau das representações utilizadas em aula que podem muitas vezes fazer o trabalho de pensar pelo aluno (quando por exemplo professores resolvem as contas pelos alunos ou os fazem memorizar fórmulas ao invés de compreenderem seu significado)
- Análise dos erros dos estudantes e de abordagens não convencionais de pensamento: avalia-se o grau em que os professores tentam identificar durante as aulas as fontes de mal-entendido e de erros dos estudantes e se as usam como base para suas explicações
- Seleção de tarefas curriculares: cabe identificar o modo pelo qual os professores usam o currículo (selecionando ou deixando tópicos de fora, colocando mais ênfase na memorização de dados ou em práticas, etc)

As evidências coletadas por Charalambous (2016) através de 65 itens distribuídos em 4 formulários mostram que os elementos diferenciadores entre os professores foram i) suas distintas capacidades de darem explicações e ii) seu uso de representações. Os professores avaliados foram classificados em grupos diferentes de acordo com suas experiências de aprendizado e de formação e resultados foram gerados através do uso da TRI (teoria da resposta ao item). Apesar da grande maioria dos testes serem destinados a professores na ativa, é possível também que eles sejam utilizados como exames de entrada na profissão (como demonstração de conhecimento mínimo, como discutem Goodman, Arbona e Rameriz, 2008).

O conhecimento dos professores pode ser testado por portais online, como feito pela Associação Nacional de Professores de Ciências dos Estados Unidos (em inglês, “National Science Teachers Association, NSTA), nos quais instrumentos específicos são oferecidos

aos professores para que possam saber seu nível de conhecimento, tais como o 'PD-Indexer' (do inglês, 'Professional Development') no qual os professores têm uma 'experiência de diagnóstico' que lhes oferece sugestões de treinamento e materiais para aqueles tópicos nos quais eles são menos proficientes. Os professores não são assim simplesmente 'testados' mas encorajados a pensar o seu desenvolvimento profissional com o apoio de recursos online (<http://learningcenter.nsta.org/indexer/>). Esse tipo de instrumento pode ser importante para a escalabilidade de inovações pedagógicas. Nem todos os recursos são grátis, mas um bom número deles, ao redor de 1.850 com 62.000 professores participantes, de acordo com o relatado por Byers et al (2011).

### **2.2.11. Modelos de Auto eficácia**

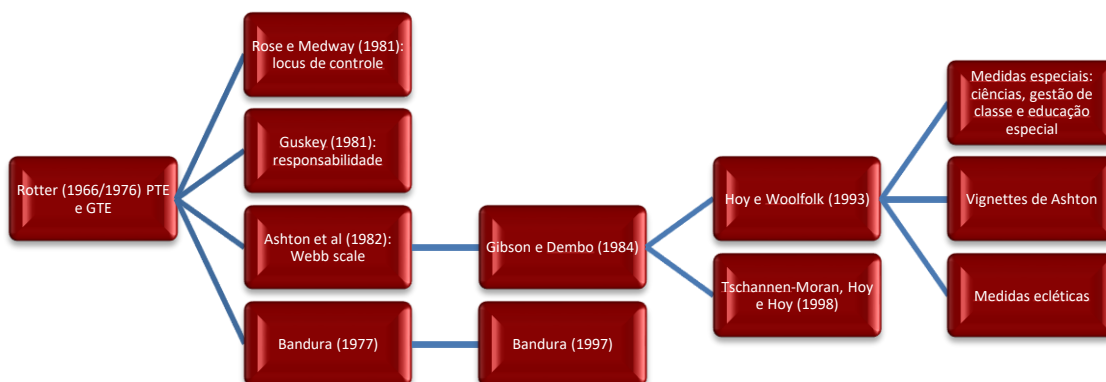
Entende-se como auto eficácia a crença que um professor tem em controlar o contexto de suas ações afetando o desempenho dos estudantes, tal como originalmente definido por Rotter (1966). Desse modo, professores com um alto nível de eficácia são aqueles que acreditam que eles podem influenciar (no seu limite controlar) o desempenho e motivação de seus estudantes. Bandura (1977), e depois colegas (como pode ser visto no diagrama abaixo), mostraram como essa crença envolve um processo cognitivo no qual as pessoas constroem uma visão sobre suas próprias capacidades de transformação dos estudantes para um dado nível de desempenho. O resultado direto disso é que professores com uma alta auto eficácia tendem a ser mais esforçados, resilientes perante as dificuldades e menos propensos ao estresse e depressões. Amorim-Neto et al (2018) mostram como a auto eficácia pode prever um comportamento empreendedor entre os professores. Medidas de auto eficácia dos professores passam assim por 'lentes psicológicas' de como eles/as encaram suas qualidades e os desafios que têm pela frente.

Vários instrumentos de auto eficácia dos professores foram desenvolvidos ao longo do tempo, a grande parte deles (usados pelos autores no quadro abaixo) como uma sequência de afirmações sobre as quais os professores deveriam indicar o seu grau de concordância. A escala mais conhecida, promovida por Bandura (1977) e (1997) resultou em 30 itens em uma escala Likert de 9 pontos organizada ao redor de 7 dimensões, quais sejam, i) influência sobre a tomada de decisão, ii) influência sobre os recursos das escolas, iii) eficácia de ensino, iv) eficácia na disciplina, v) envolvimento dos pais, vi) envolvimento da

comunidade e vii) criação de um clima escolar positivo. Depois de Bandura, um dos modelos mais estabelecidos na literatura é o de Tschannen-Moran, Hoy e Hoy (1998) que propõem um modelo integrado no qual a eficácia dos professores é mais contexto-específica. As circunstâncias de matérias ou turmas ou lugares diferentes podem afetar a auto eficácia dos professores.

Evidências produzidas para a Turquia (Alçam e Olgan, 2017) mostram como as famílias podem influir na auto eficácia dos pais e como simples cursos de formação para os pais podem não ser muito efetivos.

**Diagrama 7 – Evolução dos Modelos de Auto eficácia**



Desse modo sugerem que existem cinco fontes de auto eficácia derivadas do: i) senso de domínio das experiências prévias, dependente das experiências consideradas exitosas pelos professores; ii) nível dos sentimentos de excitação emocional e fisiológica com a experiência docente; iii) sentido de experiências imaginadas, olhando outros professores ou como aparecem na mídia ou a partir de modelos de professores bem sucedidos; iv)



convencimento verbal sobre a profissão em seminários ou em cursos específicos, convencendo os professores a inovarem e tentarem novas estratégias pedagógicas e v) do processo cognitivo dos professores no processamento de todas essas informações.

Os questionários utilizados na literatura são normalmente disponíveis como podemos ver no item 14 do Anexo desse trabalho onde apresentamos a escala de Hoy e Woolfolk de Eficácia do Professor. É interessante notar que a auto eficácia é uma variável fundamental de processo, não necessariamente de resultado. Isto é, um professor auto eficaz é uma condição necessária, mas não suficiente para um ensino de qualidade. Pode bem ser, como sugerem Goddard, Hoy e Hoy (2000) que a auto eficácia individual seja influenciada pela auto eficácia coletiva.

### **2.2.12. Modelos Mistos**

Uma conclusão importante dos relatórios do projeto MET da Fundação Gates é que “todos os instrumentos de avaliação testados foram associados positivamente com ganhos de desempenho dos alunos” (Kane et al, 2012: 6). No entanto, um segundo resultado importante foi de que a confiabilidade (“reliability” no inglês) na avaliação da prática dos professores necessita com que se trabalhe com escores médios feitos sobre diversas observações. O problema é que a nota dos professores pode variar consideravelmente de lição à lição, assim como de observador para observador. Eles, no entanto, descobriram, como seu terceiro resultado, de que a combinação de escores de observação com evidência de ganhos de desempenho dos alunos e feedback dos alunos melhora o poder preditivo e prática dessas medidas.

Em outras palavras, o MET recomenda que uma estratégia mais efetiva e confiável seja não a de ficar examinando em detalhe todos os instrumentos para decidir qual o melhor, mas ao invés focar em métodos de combinação de instrumentos alternativos. Desse modo,

a combinação de medidas leva a um maior poder preditivo, maior confiabilidade e mais estabilidade das medidas.

Outro ponto é que essas medidas combinadas são capazes de identificar professores com maiores ganhos em testes padronizados. Há uma discussão interessante baseada na evidência de que os professores que foram melhores nessa medida combinada tiveram estudantes mais vinculados emocionalmente a eles ou não.

O ponto chave dessa discussão é sobre a garantia de qualidade nos diferentes modelos de certificação. Evidências empíricas como a de que os professores devem ser observados por mais de uma lição e de que se tem mais ganho estatístico sendo observado por um observador diferente, fazem parte de um estoque de lições importantes geradas por essa literatura.

Não existe uma fórmula específica para a combinação de instrumentos variados, mas o ideal é articular diferentes espaços informacionais que tragam uma riqueza maior de informações. Assim, ao invés de combinar diferentes modelos de observação de classe parece mais produtivo combinar um modelo de observação de classe com outro modelo de valor-adicionado ou com um modelo de percepção de professores ou diretores. Modelos de práticas reflexivas parecem ampliar a fronteira dos instrumentos de medida, assim como modelos de auto eficácia parecem estar estritamente relacionados com questões de implementação de uma ampla gama de modelos pedagógicos.

### **2.3 Comparação entre os diferentes instrumentos**

As tabelas abaixo procuram comparar os diferentes instrumentos discutidos acima, primeiro em termos de sua visão pedagógica e estrutura das dimensões apresentadas, depois em termos de um conjunto de dimensões desejadas, resultantes da intersecção de várias dimensões comuns, e por fim em termos de seus pontos positivos e negativos. A ideia é que possamos examinar os principais atributos desses instrumentos e metodologias mantendo em perspectiva suas limitações.

### **Tabela 3 – Instrumentos de observação, visão pedagógica e dimensões**

<b>Instrumentos de observação</b>	<b>Visão pedagógica</b>	<b>Dimensões e sub-dimensões</b>
Valor-Adicionado	Foco em resultados. “professor bom é aquele que tem alunos com notas boas”	Quaisquer testes padronizados podem ser utilizados desde que forneçam uma base homogênea que possibilite comparação. Em geral trata somente de dimensões cognitivas (alfabetização, matemática)
CLASS (modelo de observação de aula)	Centralidade da aula e da interação entre professores e estudantes. “professor bom é aquele que dá uma boa aula”	<ol style="list-style-type: none"> <li>1. Apoio emocional (clima positivo, clima negativo, sensibilidade do professor, consideração pela perspectiva dos estudantes)</li> <li>2. Apoio ao aprendizado (qualidade do feedback, desenvolvimento de conceitos, modelagem da linguagem)</li> <li>3. Organização de classe (gestão de comportamento, produtividade, formato de aprendizado de instruções)</li> </ol>
FFT – Framework for Teaching de Danielson (modelo de observação de aula)	Visão construtivista do aprendizado. “professor bom é aquele que cumpre suas funções de professor”	<ol style="list-style-type: none"> <li>1. Planejamento e preparação</li> <li>2. Ambiente de classe</li> <li>3. Instrução</li> <li>4. Responsabilidades profissionais</li> </ol>

<p>PLATO (modelo de observação de aula)</p>	<p>Modelo pedagógico de práticas guiadas. “professor bom é aquele que segue um conjunto de práticas pedagógicas”</p>	<ol style="list-style-type: none"> <li>1. Propósito</li> <li>2. Desafio intelectual</li> <li>3. Representação de conteúdo</li> <li>4. Conexões anteriores/conhecimento pessoal</li> <li>5. Modelos e Modelagem</li> <li>6. Estratégia explícita de ensino</li> <li>7. Prática guiada</li> <li>8. Discurso de classe</li> <li>9. Feedback (nem sempre presente)</li> <li>10. Adaptações para o inglês (nem sempre presente)</li> </ol>
<p>MQI/Qualidade da Instrução Matemática (modelo de observação de aula)</p>	<p>Pedagogia focada nas especificidades dos conteúdos e pedagogia da matemática. “professor bom conhece sua disciplina e tem conhecimento pedagógico de como aplica-la”</p>	<ol style="list-style-type: none"> <li>1. Erros e imprecisão</li> <li>2. Riqueza da matemática</li> <li>3. Trabalho com os estudantes</li> <li>4. Trabalho de classe conectado com a matemática</li> <li>5. Práticas com os estudantes alinhadas aos conteúdos básicos</li> </ol>
<p>UTOP para matemática e ciências (modelo de observação de aula)</p>	<p>Pedagogia focada nas especificidades dos conteúdos das ciências e matemática. “professor bom</p>	<ol style="list-style-type: none"> <li>1. Ambiente de classe</li> <li>2. Estrutura da lição</li> <li>3. Implementação</li> <li>4. Conteúdos de Ciências e Matemática</li> </ol>

	sabe bem os conteúdos de suas disciplinas e opera em um ambiente estruturado”	
Survey TRIPOD	Modelo pedagógico estruturado com base em um diálogo/feedback dos alunos. “professor bom é aquele que sabe trabalhar em harmonia com os seus alunos”	<ol style="list-style-type: none"> <li>1. Cuidar/se importar</li> <li>2. Esclarecer</li> <li>3. Cativar</li> <li>4. Consolidar</li> <li>5. Reunir</li> <li>6. Desafiar</li> <li>7. Gerir as classes</li> </ol>
Survey de Opinião de professores (por exemplo o modelo de Stronge)	Foco nas práticas e entendimentos dos professores. “professor bom faz muitas coisas fora de sala-de-aula também”	<ol style="list-style-type: none"> <li>1. Características pessoais</li> <li>2. Gestão de classe</li> <li>3. Planejamento dos materiais</li> <li>4. Instrução</li> <li>5. Monitoramento do progresso dos estudantes</li> </ol>
Survey de Opinião dos diretores (por exemplo o modelo de Jacob e Lefgren)	Centrado em aspectos de gestão do ensino sob a ótica dos diretores. “professor bom é reconhecido como tal pelos seus diretores”	<p>Não há uma lista fixa de dimensões, mas em alguns surveys aparecem:</p> <ol style="list-style-type: none"> <li>1. Dedicção dos professores</li> <li>2. Ética do trabalho</li> <li>3. Organização dos professores</li> <li>4. Gestão de classe</li> <li>5. Modelo pessoal</li> <li>6. Satisfação dos pais com os professores</li> </ol>

INFORM	Ênfase no componente reflexivo. “professor bom é aquele que consegue refletir sobre suas práticas”	Não existem dimensões estabelecidas, mas o protocolo do ‘debriefing’ para o estímulo à discussão e reflexão dos eventos em sala-de-aula
Análise de Portfolio de professores	Pedagogia privilegia registros objetivos das práticas dos professores. “professor bom deixa registros de suas habilidades ao longo de sua trajetória docente”	Não existem dimensões estabelecidas.
Modelos de Simulação	Pedagogia das práticas em sala-de-aula. “professor bom é aquele que executa ações pedagógicas boas”	Não existem dimensões estabelecidas, mas uma análise de padrões de resposta. Amador (2017) por exemplo utiliza as dimensões: <ol style="list-style-type: none"> <li>1. Foco</li> <li>2. Expressões</li> <li>3. Temporalidade</li> <li>4. Padrão discursivo</li> </ol>
Modelos de Práticas Reflexivas	Foco pedagógico no processo de reflexão e internalização das novas	Não existem dimensões estabelecidas.

	práticas educacionais. “professor bom é aquele que reflete e internaliza boas práticas”	
Análise do Conhecimento dos professores	Modelo pedagógico baseado na verificação do conhecimento dos professores. “professor bom conhece bem sua matéria”	Não existem dimensões estabelecidas, mas quase sempre são contexto-específicas vinculadas a disciplinas particulares. No modelo de Charalambous (2016), ele utiliza: <ol style="list-style-type: none"> <li>1. Provisão e avaliação de explicações</li> <li>2. Seleção e uso de representações</li> <li>3. Análise dos erros dos estudantes</li> <li>4. Seleção de tarefas curriculares</li> </ol>
Modelos de Auto-eficácia	Modelo pedagógico focado nas características psicológicas dos professores. “professor bom acredita que pode afetar o desempenho dos seus alunos”	<ol style="list-style-type: none"> <li>1. Influência sobre a tomada de decisão</li> <li>2. Influência sobre os recursos das escolas</li> <li>3. Eficácia de ensino</li> <li>4. Eficácia na disciplina</li> <li>5. Envolvimento dos pais</li> <li>6. Envolvimento da comunidade</li> <li>7. Criação de um clima escolar positivo.</li> </ol>
Modelos Mistos	Não há um modelo pedagógico	Não há nenhuma seleção particular de dimensões ou mesmo de modelos a serem combinados.

	específico, mas critérios de previsão lastreados em medidas de valor- adicionado. “um professor bom é bom sob vários aspectos e instrumentos de avaliação”	
--	--	--

Diferentes instrumentos elegem diferentes prioridades segundo sua visão pedagógica do que significa ser um professor de qualidade. Mesmo considerando que todas essas medidas são multidimensionais surge o problema de escolha de meta-dimensões ou meta-espacos informacionais que retratem não somente o papel e processo pedagógico seguidos pelos professores mas detalhes específicos dos contextos nos quais os professores trabalham. Por isso não existe uma fórmula específica para a escolha de um modelo misto de instrumentos e indicadores. No entanto, é possível ver que as dimensões presentes nos diferentes instrumentos traçam aspectos complementares e únicos do processo docente (do conhecimento dos professores a diferentes características de seu trabalho em sala-de-aula, passando pelo que faz fora da sala de aula). Desse modo, o primeiro passo na formulação de um instrumento composto seria a combinação de instrumentos que desse a ‘melhor história’ do trabalho de um professor. É claro que outros critérios devem ser levados em conta na classificação dos instrumentos, como serão discutidos abaixo, mas o ponto de partida é pedagógico-conceitual referente a trajetória de trabalho dos professores.

Cabe notar que todos indicadores compostos enfrentam um trade-off entre foco e extensão, aplicável a questão dos modelos mistos examinados aqui. O que significa isso? Que um modelo misto que incorpore muitos instrumentos diferentes pode ter uma ampla extensão, mas perde foco, enquanto um modelo misto concentrado em poucos instrumentos tem



muito foco, mas uma extensão limitada. Assim, não há fórmula perfeita, depende do contexto que se queira avaliar e dos *stakeholders* envolvidos.

Quando comparamos mais diretamente na tabela abaixo os diferentes instrumentos em relação a um conjunto de dimensões desejadas vemos que não existe instrumento completo. Por outro lado, o potencial de complementaridade entre os modelos é alto, dado que modelos de valor-agregado, conhecimento, de simulação e portfolio são objetivos, modelos de observação de classe e reflexão buscam intersubjetividade, e surveys são completamente subjetivos. Há assim um critério possível em relação ao grau de objetividade/subjetividade dos instrumentos que pode ensejar a busca por complementaridade. Mas não existe uma regra. Do mesmo modo pode haver uma complementação em relação aos diferentes *stakeholders* envolvidos, como dito, sendo desejável escutar alunos, professores, pais, diretores e mesmo famílias.

**Tabela 4 – Comparação entre Instrumentos de Avaliação e dimensões desejadas**

Modelos/Dimensões	Impacto das notas dos alunos	Apoio emocional	Organização de classe	Apoio ao aprendizado	Tomada de decisão dos professores	Uso de linguagem em direcionada	Uso de tecnologia	Liderança, valores e práticas pessoais compartilhadas	Autonomia reflexiva dos professores
Modelos de Valor-Adicionado ("Value-Added")	X								
Modelos de Observação em Classe		X	X	X		X		X	

e - CLAS S									
FFT – “Fram ework for Teachi ng” por Daniel son			X	X	X	X		X	
PLAT O			X	X	X	X			
MQI			X	X	X	X			
UTOP			X	X	X	X			X
Surve ys de opiniã o de estuda ntes (e.g. Tripod Surve y)		X	X	X					
Surve ys de opiniã o de profes sores		X			X				
Surve ys de opiniã o de diretor es	X	X	X		X				
Model o INFO RM de observ ações de campo					X	X	X		
Análise de <i>portfoli</i>			X	X					

o dos profes sores									
Model os de simula ção			X	X	X				
Model os de prática s reflexi vas							X	X	X
Análís e do conhe ciment o dos profes sores				X	X	X			
Model os de Auto- eficáci a		X	X	X	X			X	X
Model os mistos	X	X	X	X	X	X	X	X	X

A tabela abaixo nos mostra que não há instrumento sem limitações. Mas que diferentes instrumentos possuem diferentes justificativas que podem ser exploradas de acordo com o propósito avaliativo que se tenha definido. Por exemplo, modelos de avaliação mais específicos ao objeto de estudo como o MQI e UTOP devem ser ajustados ao nível específico de conhecimento matemático ou de ciências que se esteja trabalhando. Do mesmo modo, deve-se levar em conta que novos instrumentos pedagógicos utilizados por computador levam a uma desmaterialização dos portfólios dos professores e que isso pode avaliar com menos precisão o trabalho feito pelos mesmos fora da sala-de-aula. O Anexo I fornece uma lista de instrumentos traduzidos e materiais que podem ser utilizados para a elaboração de instrumentos mistos de avaliação de professores.

#### **Tabela 5 - prós e contras de cada método**

<b>Modelos/Dimensões</b>	<b>Prós</b>	<b>Contras</b>
Modelos de Valor-Adicionado (“Value-Added”)	Fornecem medidas objetivas, ideais para gestores	Problemas de estimação, endogeneidade, não fornecem medidas diretas de avaliação de professores
Modelos de Observação em Classe como CLASS, FFT, PLATO	Fornecem sugestões diretas para professores melhorarem suas ações e práticas	Não cobrem todas as dimensões. Não são disponíveis livremente.
MQI e UTOP	Refletem especificidades das disciplinas que precisam ser avaliadas de acordo com suas idiossincrasias pedagógicas	São limitados em escopo.
Surveys de opinião de estudantes (e.g. Tripod Survey)	Úteis para o universo que pode ser apreendido pelos alunos; livremente disponíveis	Não tem alcance pedagógico
Surveys de opinião de professores	Mais fácil de serem entendidos e assimilados pelos próprios professores, bom alcance	Sujeitos a distorções ou posições estratégicas dos professores
Surveys de opinião de diretores	Evidência positiva sobre sua validade e	Nem sempre uma perspectiva top-down será

	confiabilidade, com boa cobertura	bem aceita no momento de implementação
Modelo INFORM de observações de campo	Protocolos e rubricas disponíveis e proximidade com CLASS	Alcance limitado
Análise de <i>portfólio</i> dos professores	Medida objetiva; evidencia concreta de preparação e planeamento de classe	Alcance limitado
Modelos de simulação	Possibilitam avaliação de como os professores reagiriam em certas situações com custo muito baixo	Podem não ser levados a sério pelos professores
Modelos de práticas reflexivas	Facilitam a internalização e adoção dos protocolos	Deixam de lado medidas objetivas
Análise do conhecimento dos professores	Permitem medidas objetivas	Incompleto, pois não é somente o conhecimento da matéria que define a qualidade de um bom professor
Medidas de Auto eficácia	Medidas de impacto de curto-prazo	Especificação de contextos específicos
Modelos mistos	Tem o potencial de reunir todos os benefícios dos métodos escolhidos	O portfólio escolhido pode também trazer as limitações dos métodos contemplados

De modo geral essa análise mostra que existe uma diversidade de dimensões contempladas. O fato é que diferentes avaliações servem a diferentes propósitos e é fundamental que se esclareça finalidade e o modelo pedagógico que contempla antes da definição de um modelo de avaliação de professores. A exceção cabe aos modelos mistos que são mais inclusivos. Mas isso depende é claro do que seja a composição particular de cada um dos tipos de modelo misto.

### **3. Critérios Técnicos para Avaliação de Testes**

A maior parte dos artigos examinados, como Hamilton, Stecher e Klein (2002) trabalha com critérios de precisão, validade e confiabilidade. Mas trabalhos específicos como os de Le e Klein (2002) e Amrein-Beardsley, Polasky e Holloway-Libell, (2016) tratam também com critérios menos utilizados como o de justiça (“fairness”), sugerindo que o igual tratamento de todos envolvidos é um aspecto a ser considerado na implementação dos processos de avaliação. Outro critério relevante e pouco usado é o critério de transparência associada a ‘accountability’ (Deluca, 2012). Também encontramos na literatura (Hallinger, Heck e Murphy, 2014) menção ao critério de ‘estabilidade’ da medida de avaliação dos professores. Em particular, critica-se as medidas de valor-agregado por gerarem padrões instáveis de resultados ao longo do tempo, como discute Darling-Hammond et al (2012). Mais recentemente, Hawley, Bovaird e Wu (2017) mostraram como medidas baseadas em modelos multivariados produzem resultados mais estáveis em comparação com medidas unidimensionais. É possível que esses critérios sejam menos utilizados por envolverem aspectos normativos e sociais que podem ser abertos à discussão, diferentemente dos outros aspectos discutidos abaixo que são mais técnicos.

Um ponto de partida importante é o reconhecimento de que não existe um teste que possa medir tudo. Por isso sempre é necessária uma análise prévia sobre quais são os principais domínios de interesse. Também é verdade que devemos ter em conta não apenas o desenho e construção dos instrumentos, testes e indicadores, mas como os escores são utilizados. No que segue listamos os principais critérios e seus significados:

1. **Viés:** essa é uma preocupação que está presente em quase todos os instrumentos de avaliação, mesmo que tenha origem em problemas diferentes. Nos modelos de valor adicionado as principais fontes de viés estão relacionadas a problemas de seleção ou especificação do modelo. Nos modelos de observação de classe, quase sempre o problema é a subjetividade dos avaliadores. Nos modelos de percepção dos alunos também é usualmente uma questão relacionada a sua subjetividade. No entanto, também é verdade que nos relatórios de observação de classe há o problema do que de fato se está avaliando. Muitas vezes o que se avalia não é a contribuição isolada dos professores, mas um conjunto de contribuições ou mesmo a interação entre os professores e os alunos. De fato, o desenvolvimento das competências dos alunos é quase sempre o resultante de um processo interativo onde não apenas professores e escolas têm uma contribuição fundamental, mas os próprios alunos;
2. **Confiabilidade:** o problema da falta de confiabilidade surge porque muitas vezes as avaliações precisam ser feitas a partir de uma amostra relativamente limitada de observações dos professores. A pergunta chave para definir esse critério é: “se fizéssemos essa mesma avaliação sob outras circunstâncias, por exemplo, vindo outro dia, ou com outro avaliador, ou se os estudantes fizessem a prova em outro dia, teríamos a mesma informação?” Há uma discussão importante sobre o anúncio ou não de que as observações serão feitas aos professores (se sua preparação influencia ou não os resultados). Há também o tema de como combinar tipos diferentes de avaliadores e de como treiná-los. No entanto, medir a correlação entre diferentes avaliações e cruza-las parece gerar bons resultados. Podemos dizer que os principais fatores que afetam a confiabilidade de um resultado é a) amostragem, b) fatores transitórios, como se um estudante particular fez uma prova com boa saúde ou não, se a sala tinha luz adequada, o dia do teste, etc, c) grau de acordo entre os avaliadores (“raters”) e d) no caso dos modelos de valor-agregado o que afeta é a duração e formato dos testes. É possível, no entanto quantificar a confiabilidade de um teste. Existem três métodos principais: dando versões diferentes de testes aos estudantes, testando-retestando em ocasiões diferentes e usando métodos de consistência interna, que olham para o grau de correlação entre questões diferentes no mesmo modelo de teste ou entre itens dentro de cada parte.

Essa parte pode ser expandida, tratando dos coeficientes de confiabilidade e do uso do erro padrão de mensuração, tal como descrito por Le e Klein (2002), Kane et al (2012) e Ho e Kane (2013).

Um ponto essencial nessa discussão é que confiabilidade é um critério estratégico pois estamos interessados não somente em que pessoas altamente qualificadas possam avaliar professores, mas que para que esses instrumentos de avaliação sejam úteis para as escolas deve-se considerar que eles sejam transferíveis para pessoas com um nível mais de *expertise*. Não se trata apenas de uma questão de escalabilidade, mas de grau de *transmissividade* da aplicabilidade dos instrumentos.

3. **Validade:** a validade é considerada atualmente um conceito unificado no qual tipos diferentes de evidência contribuem para um julgamento composto de validade. Como destacam Lee e Klein (2002), podem, no entanto, haverem circunstâncias nas quais é feita uma referência a algum tipo específico de validade, como a validade da instrução. Para autores como Papay (2012), essa propriedade é a mais importante de todas, mas por outro lado é a mais difícil de avaliar. Isso acontece porque a validade de uma avaliação é uma propriedade de uma conclusão que se espera tirar da medida, não uma propriedade da medida por si mesma. Por essa razão podemos dizer que avaliar a validade do aprendizado dos estudantes é uma coisa enquanto avaliar o aumento das notas de suas provas é outra. Questões como o nível de (sub) representação do *constructo* que se quer avaliar (quando se falha em capturar o domínio desejado) ou como a irrelevância dos conteúdos são chave para o estabelecimento da validade. Por isso nem sempre a evidência proveniente do teste de conteúdos pode ser considerada válida.

Diferentes testes podem ser vistos como expressões parciais de padrões de relações entre medidas diferentes. Assim, um teste que se propõe a medir o conhecimento de operações básicas com números reais deve ter uma forte relação com outro teste



que tenha o mesmo objetivo. Portanto *evidências de discriminação por convergência* têm tido um papel importante em avaliações recentes (Yen e Ferrara, 1997 e Hambleton et al, 2000). Avaliações feitas intra-medidas, testando a correlação entre subconjuntos de escalas ou itens de teste individuais, podem ter um efeito semelhante de validade.

O ponto geral do uso de critérios, sejam esses ou outros, é que testes e/ou instrumentos podem conter vieses e que por isso devem ser eles mesmo objetos de avaliação. Algumas vezes isso pode ser complicado pois essas avaliações devem ser feitas em relação aos propósitos que servem. Isso quer dizer que alguns critérios que sirvam para qualificar algumas avaliações podem não servir para outras (evidências coletadas pela *American Educational Research Association* e *Natural Council on Measurement in Education* corroboram esse ponto).

Uma boa utilização desses critérios deve levar a um conjunto de questões, como listadas no quadro abaixo, que deve ser contemplado na elaboração de um novo instrumento ou mesmo na aplicação de um dos instrumentos conhecidos:

#### **Quadro 1 – Desenho das Observações: calculando os vieses**

1. Avaliadores diretores vs avaliadores professores
2. Professores da mesma escola vs professores de outras escolas
3. Lições selecionadas por professores vs sem aviso prévio
4. Avaliadores certificados de graus diferentes
5. Tempo curto vs tempo longo de avaliação
6. Quantas classes pedir para cada avaliador (considerar atrito dos avaliadores)
7. Alocação de classes/vídeos entre os avaliadores: estruturado vs randomizado
8. Avaliadores internos vs avaliadores externos
9. Escolha da abordagem pedagógica (“framework for teaching”)
10. Construção das rubricas
11. Tempo de duração dos vídeos

12. Matriz de componentes de variância e confiabilidade

13. PS: o uso da Teoria da Generabilidade de Brennan (2001) pode oferecer perspectivas interessantes de cruzamento dessas informações. Trata-se aqui de estabelecer uma ferramenta, como uma matriz, que possibilite a comparação entre diferentes resultados provenientes de métodos de avaliação diferente, gerando possíveis explicações e elementos reflexivos para uma análise mais completa.

Há toda uma literatura, como Kordts-Freudinger e Geithner (2013) que mostra evidências de como as taxas de respostas a surveys são mais elevadas quando aplicadas em classe do que quando deixadas para serem aplicadas depois. Interessante notar também que as próprias notas parecem ser mais negativas do que aquelas dadas em classe. Por fim, aqueles que fizeram fora de classe escreveram comentários negativos mais longos, independente de se usar questionários online ou em papel.

#### **4. Implicações Práticas (para políticas públicas)**

Alguns autores da corrente de valor-adicionado, como Hanushek e Rivkin (2010) sugerem que um uso possível para a informação sobre a heterogeneidade dos professores é a eliminação dos piores 6-10% que teria segundo eles um efeito dramático no desempenho médio dos alunos. Mas essas medidas podem ser usadas também para incentivos, oferta de emprego, promoções ou mesmo para decisões de alocação de professores. Claro, o ponto principal nesses usos é o poder de atribuição dos modelos ao trabalho do professor (Austin et al, 2015). No entanto, não podemos esquecer também de considerações outras como justiça. O uso isolado de medidas de valor-agregado pode ser insuficiente para fazer justiça a um professor se existem fatores não-observáveis que influenciem o baixo desempenho dos alunos. Nesse caso a sugestão é o uso de medidas complementares ou mesmo de um modelo misto de instrumentos. Fernandez et al (2018) sugerem o uso de políticas de avaliação de professores que incluam famílias. Evidências baseadas na aplicação de um projeto piloto em 12 distritos de Connecticut (Estados Unidos) mostrou que resultados dependem muito do modo pelo qual a participação dos pais é promovida: como stakeholders legítimos ou na periferia das relações com a escola. Essa é uma área no entanto que merece mais atenção.

Distorções por incentivos também são possíveis. Aguilar e Richerme (2014) argumentam como o uso de indicadores de desempenho pode levar a várias consequências indesejáveis afetando até mesmo a imparcialidade do trabalho científico (afetando até mesmo o tipo de linguagem utilizada). Seja como for, um outro problema potencial é focar somente na variação dentro das escolas (o que é feito naturalmente pela dificuldade de explicarmos variações entre escolas) ignorando que em média algumas escolas têm conjuntos melhores de professores do que outras.

No balanço, medidas de valor-agregado parecem pouco úteis como geradoras de sugestões e recomendações aos professores (que podem ser diferentes entre escolas públicas e privadas, ver Bolossi e Hernandez, 2016), o que significa que possuem um potencial muito limitado como instrumento de desenvolvimento dos professores (Bonner et al, 2018). No entanto, conferem uma boa base objetiva de avaliação dos professores, mesmo que limitada àquelas matérias que são mais passíveis a padronização em testes. Se o objetivo de política pública é a melhoria de formação dos professores, fica claro que modelos de observação de classe têm um potencial muito maior. E, tratando-se de um conhecimento mais específico, na área de ciências ou matemática, deve-se considerar modelos de observação de classe que considerem essas idiosincrasias.

Uma preocupação com a implementação dessas estratégias formativas deve contemplar elementos de 'debriefing' como vimos em alguns modelos de avaliação bem como de 'práticas reflexivas' que levam a uma maior e melhor internalização dos conteúdos pedagógicos. As evidências coletadas, apresentadas acima, corroboram o uso desses elementos.

Diferentes modelos de avaliação, baseados em distintos modelos pedagógicos, têm implicações diferenciadas sobre os melhores modos de capacitar os professores. Em linhas gerais é possível sugerir um conjunto de práticas derivadas da discussão acima, que dever tornar mais eficiente a capacitação dos professores, tais como:

1. Professores devem ser capacitados em gestão de classe. Não basta transferir conhecimento ou informação aos professores. Eles devem ser capazes de internalizar conhecimentos práticos, concretos que os ajudem a navegar melhor entre seus alunos e contextos difíceis;
2. Professores devem ser capacitados em atributos psicoemocionais que são indispensáveis, segundo as evidências encontradas, para um ensino eficiente;
3. Professores devem ser capacitados através de instruções concretas que possam ser entendidas como parte do seu dia-a-dia; capacitações em teorias abstratas não tem grande potencial de mudança do comportamento concreto dos professores;
4. Professores devem ser capacitados para lidar com o imprevisto e o inesperado, pois isso é um padrão identificado como vigente em muitas salas-de-aula;
5. Pela mesma medida, professores são melhor capacitados dentro do seu ambiente de trabalho (escolas) ao invés de o serem em outros ambientes que aumentem a dificuldade de transferência do seu conhecimento (pois esse deve ser prático, levando em consideração as circunstâncias de aplicação)
6. Capacitações devem respeitar o tempo de interpretação, apropriação e transformação dos professores; não se deve esperar que os professores sejam apenas reprodutores de fórmulas pedagógicas definidas em gabinetes. Muitas vezes isso significa também permitir um tempo de apropriação coletiva;
7. As evidências coletadas sugerem que modelos mistos de avaliação parecem ser os mais frutíferos pois permitem o estabelecimento de parâmetros de validade e confiabilidade que atendam diferentes propósitos. Se possível, feitos em consulta com os próprios professores e atendendo os protocolos específicos de cada instrumento específico.

Alguns modelos recentes, como o de Anderson-Levitt, van Draanen e Davis (2017) sugerem que é possível desenvolver estratégias de formação nas quais os estudantes de pedagogia possam aprender através das dissonâncias criadas entre teoria e prática e busca de coerência (por identidade ou valores universitários). Programas de residência de estudantes podem ser usados para essas formações com bons resultados.

## **5. Consulta Empírica**

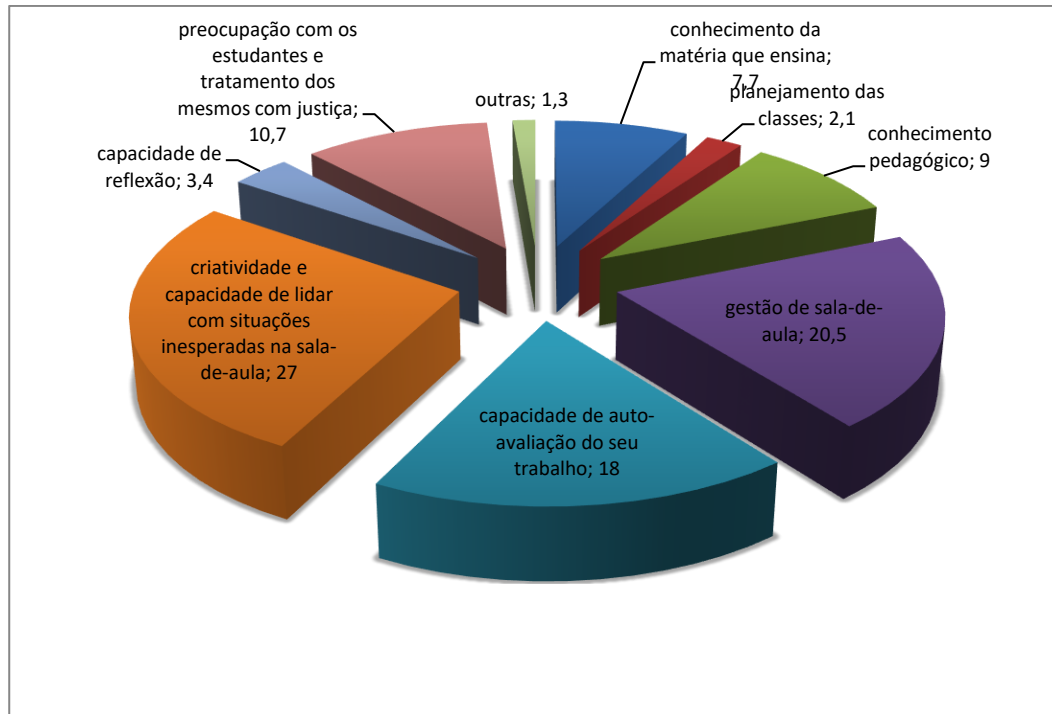
Como parte desse estudo foi realizada uma pesquisa empírica sobre o tema da avaliação de professores com gestores educacionais, gestores públicos, professores do ensino fundamental e universitário, especialistas em educação, jornalistas e pesquisadores da área educacional. O objetivo foi usar uma amostra experimental, constituir um *benchmark* de referência, sem ambição de validade amostral externa, para dialogar com alguns resultados analíticos obtidos aqui. Foram aplicados 260 questionários dos quais 236 foram aproveitados. Alguns traços característicos de nossa amostra: 60% das pessoas que responderam foram gestores públicos vinculados a secretarias municipais de educação, com uma idade média de 46 anos e com alto grau de instrução (70% com pós-graduação). A amostra foi constituída de 70% de mulheres, 20% de homens e 10% que responderam institucionalmente sem identificar gênero.

A pesquisa perguntou primeiro genericamente o que é essencial para que uma pessoa seja um bom professor ou boa professora (através de uma pergunta aberta). Dentre as respostas com mais alto nível de frequência encontramos que apenas 17% sugeriu 'uma boa formação' como um elemento importante, enquanto 25% mencionou 'comprometimento', 10% 'amor', 9.5% 'dedicação', 7% 'vontade', 6% 'vocação' e 5% 'gostar do que faz', totalizando quase 63% de fatores relacionados a aspectos psicoemocionais dos professores. Por um lado, podemos ver esse resultado como distinto dos sugeridos pela maioria dos modelos de avaliação que buscam objetividade, intersubjetividade ou introspecção, mas por outro parece oferecer uma boa ilustração de evidencia obtida para a China na qual os professores sugerem que os parâmetros mais importantes de desempenho dos próprios professores são medidos como a moralidade, ou o potencial de colaboração ou o próprio senso de identidade profissional dos mesmos (Liu et al, 2016)

Desse modo parece que, diferentemente da pesquisa teórica 'padrão' realizada aqui, que pode existir um entendimento entre muitos especialistas que fatores motivacionais estão na base da qualidade dos professores. É interessante notar, contudo, que raramente esses aspectos psicoemocionais são tratados na formação regular ou extra de professores.

Uma pergunta relacionada, mas mais específica, pediu que as pessoas definissem o que é um professor auto eficaz. Nesse caso foi dado um reconhecimento maior para o elemento de boa formação, capacitação e atualização dos professores que chegou a 36%, seguido de 15% que apontou a característica de ‘gestão de sala-de-aula’, 9.4% ‘comprometimento’, 7% ‘confiança’ e 4.3% ‘inovação’, entre outros. Uma outra pergunta, mais específica ainda, perguntou diretamente aos entrevistados quais características achavam mais importantes para definir um professor de qualidade. Os resultados são apresentados no diagrama abaixo. Encontramos um alto percentual de respostas sobre o que é um professor de qualidade em um conjunto de fatores considerados mais avançados na literatura, como a gestão de sala-de-aula, capacidade de auto avaliação e criatividade & capacidade de lidar com situações inesperadas na sala-de-aula. De algum modo essas respostas apoiam a relevância de vários estudos discutidos aqui ao fazerem parte do universo conceitual (e presumivelmente da vivência) dos entrevistados.

**Diagrama 8 – O que é um professor de qualidade?**



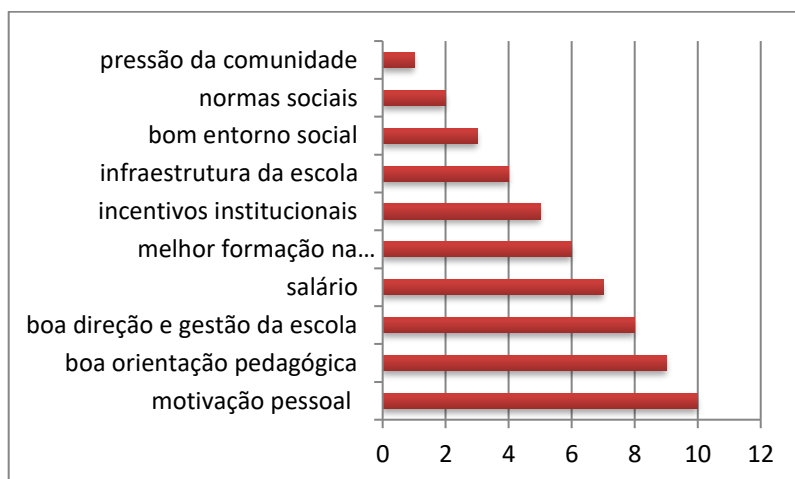
As pessoas entrevistadas também deixaram claro que consideram “a influência dos professores” o maior fator determinante das notas dos alunos, seguido da “influência das famílias” e por último da “influência da direção”. Quando abrimos esses dados vemos que

de um lado existe uma grande concentração de respostas (56.5% delas) entre as pessoas que acreditam que a influência dos professores é superior a 80%, enquanto, de outro apenas 27% acredita que as direções das escolas têm essa influência.

Parece haver um desacordo entre os entrevistados sobre o grau desejado de intervenção dos professores em sala-de-aula. De um lado encontramos 42.7% que acham que os professores devem intervir em tudo que podem em classe; de outro 49.6% acreditam que os professores devem tentar resolver tudo pelo diálogo. Enquanto isso 7.7% revelou pensar que os professores não devem intervir em nada e deixar a aula rolar segundo a vontade dos alunos.

Mas será que o trabalho dos professores depende deles mesmos ou das suas condições de trabalho? Esse é um fator complexo, discutido extensivamente na literatura de valor-agregado (com problemas sérios de endogeneidade nos modelos além de falhas de especificação dos mesmos por variáveis omitidas), mas que segundo a consulta realizada depende de uma multiplicidade de fatores (ordenados no gráfico abaixo) cujas interações geram um grau considerável de complexidade na interpretação e análise da 'função de produção educacional'.

**Diagrama 9 – ranking de fatores que afetam o desempenho dos professores**



O gráfico acima mostra uma ordenação dos fatores que foram obtidos através de respostas a uma escala Likert de 5 pontos na qual a nota mais baixa foi de 3.5 (sugerindo que mesmo o fator mais baixo era também importante). Talvez cause surpresa para alguns economistas que o salário e incentivos institucionais não estejam entre os primeiros elementos destacados pelas pessoas consultadas. De fato, os três primeiros tratam de motivação pessoal (fatores intrínsecos), orientação pedagógica e gestão da escola. Mas todos os outros fatores também são considerados relevantes, o que significa que as ações de melhoria do desempenho dos professores devem ser *ações integradas* (atuando em muitas frentes simultaneamente ou sequencialmente). Em particular, a avaliação de professores pode contribuir para uma melhoria dos princípios pedagógicos efetivamente usados assim como para uma mudança dos currículos universitários.

Há um consenso na amostra que as avaliações são um importante instrumento de melhoria da qualidade dos professores e da qualidade do ensino, sinalizado por 85.4% das pessoas consultadas que acham que os professores do ensino fundamental deveriam ser avaliados em bases regulares (semestrais ou anuais). É relevante notar que não há diferença significativa na visão de gestores e professores do ensino fundamental nesse quesito. Para 64% dos entrevistados a maior vantagem nesse tipo de avaliação é poder contar com uma medida de melhoria de qualidade do ensino. Um percentual de 12% acha que as avaliações podem ser utilizadas como medida de promoção (incentivos) aos professores. Por outro lado, 10% acha que as avaliações possam ter nenhuma utilidade. Um mesmo percentual de 10% acredita que as avaliações podem ser usadas para a promoção de capacitação daqueles professores menos preparados.

Por fim, apresentamos uma lista de todos modelos de avaliação estudados aqui, perguntando aos entrevistados quais os modelos que eles consideram mais e menos úteis. A tabela abaixo mostra os extremos da distribuição entre aqueles que consideram os modelos particulares nada útil e o 'mais útil'. Vemos que os modelos de valor-adicionado, surveys de opinião de estudantes, professores e diretores estão entre os considerados 'nada úteis' enquanto modelos voltados à observação de classe, análise de portfólio, conhecimento e práticas reflexivas de professores estão entre os considerados mais úteis.



Nenhum, no entanto ganha dos chamados modelos mistos (onde vários desses modelos são combinados). O grau de aceitação dessas metodologias é um elemento chave para que possamos tratar não somente dos méritos técnicos dos diferentes modelos, mas também dos seus diferentes graus de aceitação entre diversos *stakeholders*, peça fundamental para sua implementação. O melhor modelo de avaliação do mundo serve muito pouco se não existirem as condições para sua implementação efetiva.

**Tabela 6 – utilidade dos modelos de avaliação**

Modelos	% dos que consideram o modelo nada útil	% dos que consideram o modelo o mais útil de todos
Valor-Adicionado	60	9
Observação de Classe	44.5	25
Surveys de opinião de estudantes	59	9
Surveys de opinião de professores	62	11
Surveys de opinião de diretores	63	12
Modelo INFORM de observações de campo	56	18
Análise de portfólio de professores	35	28
Modelos de simulação	49	20
Modelos de práticas reflexivas	39	25
Análise do conhecimento dos professores	46	26
Modelos mistos	38	33

Um alto percentual, de 80.5% das pessoas entrevistadas, considera que algum desses modelos é muito útil. Combinando essa estatística com a anterior sobre a utilidade das avaliações regulares podemos ver, para esse conjunto de pessoas entrevistadas, que há um ambiente favorável as discussões sobre a melhor forma de avaliação dos professores. Mas esse é um assunto complexo. Várias regressões foram rodadas como parte desse estudo tentando explicar uma visão favorável ou desfavorável as avaliações por parte das pessoas entrevistadas, mas os resultados foram inconclusivos. Desse modo apenas

podemos observar que há uma certa apreciação, por parte da nossa amostra experimental, dos modelos de avaliação que reflete uma tendência similar à tendência pesquisada teoricamente. O único ponto de maior discordância talvez seja em uma menor valorização das formações na geração de professores de qualidade em relação à literatura consultada.

## 6. Lições das Experiências Internacionais

A grande maioria dos países do mundo tem sistemas de avaliação de professores. Mas existe uma grande diversidade de modelos de avaliação bem como países que optaram por não avaliar seus professores (e outros em transição, como o caso de Cingapura). Os países que avaliam diferem no grau de centralização/descentralização da aplicação de exames padronizados, no grau de autonomia que conferem a gestão e à avaliação nas escolas, na periodicidade das avaliações, na existência ou não de departamentos especializados na implementação das avaliações, no uso de instrumentos diferenciados (de medidas de valor-agregado a medidas de auto avaliação), no link com a formação dos professores e uso ou não de medidas de certificação, e finalmente no estágio curricular em que implementam essas avaliações (ensino fundamental, médio ou universitário).

No que segue foram sistematizadas na tabela abaixo informações sobre sistemas de avaliação de professores de 31 países. Na sequência apresenta-se um esquema com algumas reflexões sobre características marcantes desses sistemas.

**Tabela 7 – Experiências internacionais de avaliação de professores**

País	Como funciona	Tipos de avaliação usadas	Inovações pedagógicas
Japão	Avaliações descentralizadas feitas por escolas; País e comunidades também conduzem avaliações externas	VA (valor adicionado)	Uso de trabalho em grupos em classes de matemática do ensino médio; Mais pedidos de explicação aos estudantes (entre 2007 e 2011, subiu de 28.8% para 45.5% o percentual de estudantes do 8º ano explicando o que estão estudando a seus professores de

			ciências) Fonte: Educational Policy Outlook Japan
Finlândia	Reforma educacional no final dos anos 70 estimulou professores do ensino fundamental a terem mestrado; Somente 10% dos candidatos são aceitos como professores do fundamental; Currículo na faculdade tem ênfase na prática	Ministério da Educação não tem avaliação de professores (não possuem testes padronizados); Professores são avaliados de acordo com o currículo nacional e o plano de desenvolvimento da escola	Programa de Instruções profissionalizantes (Ammattistartti) tem diminuído abandono escolar; A Rede Finlandesa (Osaava 2010-2016) para treinamento de professores apoia formação local de professores; Professores têm autonomia pedagógica. Fonte: Educational Policy Outlook Finland
Dinamarca	Tradição de autonomia das escolas; Possuem duas autoridades nacionais de avaliação: Agência Nacional para Qualidade e Supervisão (2011) e o Instituto de Avaliação da Dinamarca (1999)	VA, surveys de professores e surveys de diretores; A avaliação dos professores é voluntária e práticas como as avaliações internas das escolas são definidas localmente	Foi feita uma grande reforma da educação dos professores em 2012 baseada em desregulamentação, internacionalização e fortalecimento da conexão com as necessidades das escolas públicas. Fonte: TALIS
Estônia	Cada escola desenvolve o seu currículo individual baseado no currículo nacional, o que dá poder limitado para as avaliações padronizadas.	Inspeções externas, VA, avaliações internas regulamentadas pela 'Lei das Escolas Secundárias e Básicas'. As avaliações de professores são usadas para promoções e formações.	Foi introduzido um programa de líderes escolares (responsáveis pela gestão de recursos e implementação de currículos) em 2015 junto com um novo modelo de certificação de professores; Estas políticas foram sincronizadas com uma 'Estratégia de Aprendizado de toda vida' a ser alinhada com salários de professores até 2020. Fonte: Educational Policy Outlook Estonia e TALIS
Suécia	Avaliações padronizadas junto com avaliações escolares de processos de qualidade	Não existe um sistema de avaliação oficial de professores. Mas VA e surveys de professores são usados em vários tipos de avaliação. Falta um framework nacional coerente para integrar diferentes avaliações; Dentro do programa de treinamento de professores em matemática (Matematiklyftet) existem reuniões entre os professores para debriefing como o modelo INFORM.	Existe um programa de formação de professores (Lararlyftet –primeira fase de 2007 a 2011 e segunda fase 2012-2018) que oferece a eles cursos especializados e formação focada a leitura e matemática. Também existem programas similares para formação de líderes escolares. Fonte: OCDE
Australia	Avaliações dos professores são definidas pelo Instituto Australiano	Esquema de certificação nacional baseado no conhecimento dos professores (voluntário, recebem feedback de suas práticas) e auto-	Preocupação com a questão migratória levou a um programa de avaliação e treinamento para imigrantes promovidos

	para Ensino e Liderança (AITSL)	avaliação (promovido por recursos online livres).	pelo mesmo AITSL. Fonte: Governo da Austrália.
Coréia do Sul	Todos os professores são avaliados anualmente independentemente do seu tipo de contrato.	Surveys de professores e de estudantes/famílias. Focada na promoção e carreira dos professores. Pode levar à punição e mesmo demissão de professores.	Novo programa de avaliação introduzido foca em dar feedback as práticas dos professores. Avalia também a capacidade de gestão dos diretores; Disseminação de histórias de professores bem-sucedidos para a criação de uma cultura de respeito aos professores na sociedade. Fonte: OCDE e Banco Mundial
Irlanda	Não existe um sistema formal de avaliação de professores	Os professores são avaliados como parte de inspeções gerais nas escolas. Um retorno é dado aos professores como parte desse processo. Os principais critérios usados são: conhecimento da matéria, preparação das aulas, abordagens de acesso aos alunos, responsividade a necessidades individuais dos alunos e práticas colaborativas.	Novo documento lançado focado na avaliação: "Looking at our schools 2016: a quality framework for post-primary schools". Fonte: sites oficiais do governo da Irlanda.
Inglaterra	Avaliações são feitas pelo Escritório de Padrões na Educação (Ofsted) de caráter nacional.	Inspeções geram relatórios utilizados pelas famílias na busca de boas escolas. Informações complementadas por exames e provas nacionais. Um dos países referência para VA (valor adicionado)	Foi criada em 2000 a Escola Nacional de Liderança Escolar (NCSL) e introduzidas práticas de responsabilização e gestão nas escolas. Fonte: O Sistema de Formação de Lideranças da Inglaterra. British Council e Itaú Social.
Holanda	Avaliações descentralizadas feitas pelas escolas.	Existe um regulamento federal que determina que as escolas tenham reuniões para discutir o desempenho dos professores. Na prática isso é feito em 60% das escolas	O Sindicato dos Professores introduziu um sistema de registro e de revisão de pares ("peer review") para professores em 2011. No registro os professores podem dizer o que tem feito. Na revisão de pares os professores podem observar classes de outros professores na sua própria escola ou podem ir a uma escola diferente. Fonte: sites do Governo da Holanda.
Alemanha	O sistema de avaliação de professores é coordenado pelo Ministério da Educação.	Avaliações feitas por agências externas com critérios para avaliação dos professores tais como: planejamento de classes, promoção de ambientes de aprendizado, capacidade de motivação dos alunos, consciência de diferenças sociais e culturais, capacidade de resolução de conflitos e entendimento da profissão	Foi desenvolvido um framework de garantia de qualidade de professores em 2005 que inclui dois exames estatais e um treinamento. Fonte: Ministério da Educação da Alemanha e KMK.

		como um processo de aprendizado contínuo.	
Portugal	O sistema de avaliação de professores foi inicialmente introduzido em 2007 e é atualmente regulado pelo "Sistema Integrado de Avaliação do Desempenho da Administração Pública" (Decreto Lei 41/2012). As avaliações ocorrem em ciclos de 4 anos.	Auto-avaliação e pareceres de avaliadores internos (em geral coordenadores pedagógicos). Usa 3 instrumentos: 1) projeto docente, 2) documento de registro e 3) relatório de auto-avaliação e parecer do avaliador interno. Também são utilizadas observações de classe mandatórias para professores em estágio probatório e nos estágios 2 e 4 da carreira assim como para aqueles que querem chegar no topo da carreira e para aqueles classificados como de desempenho insuficiente. (usam no total 180 minutos divididos em ao menos dois momentos)	Avaliações empreendidas academicamente (Santiago, 2012) sugerem que a política de avaliação de professores em Portugal fornece bons parâmetros de progressão de carreira e de auto-avaliação que estimulam o desenvolvimento profissional. Fonte: documentos do Governo de Portugal.
Espanha	A "Lei Orgânica de Educação" (artigos 106 e 141) estabelecem a avaliação de professores para todos os âmbitos educativos. Mas não há nenhuma estrutura de avaliação de professores funcionando para a educação básica.	Cada comunidade autônoma é responsável pelos seus próprios instrumentos e está em um nível distinto de avaliação de professores. Enquanto a "Agencia de Evaluación Educativa" da Andalusia contempla auto-avaliação e avaliações externas, a "Comunidad Autónoma del Principado de Asturias" tem um sistema de avaliação baseado na implementação de planos pedagógicos.	O "Instituto Nacional de Evaluación Educativa" (INEE) trabalha para a produção de avaliações baseadas em testes padronizados da OCDE, como o PISA. Para isso produz o "Sistema de Indicadores de la Educación" em bases regulares. Na prática somente 10% dos estudantes espanhóis estão em escolas onde as classes são observadas por professores sênior ou diretores. Fonte: OCDE
Taiwan	Para ensinar os professores devem passar primeiro em um teste de certificação do governo central.	Avaliações anuais feitas pelas escolas que incluem itens de monitoramento como desempenho dos professores em tarefas administrativas, 'profissionalismo' até características pessoais.	Esquema de certificação para professores iniciantes. Fonte: Governo de Taiwan.
África do Sul	A cultura do apartheid (com controles de escolas) deixou uma herança contrária a avaliações e professores desconfiados.	O 'Departamento de Educação Básica' estabelece parâmetros curriculares e de exames, mas as avaliações são feitas com base em surveys de professores e auto-avaliações dentro de um 'Sistema de Gestão de Qualidade Integrada' (IQMS)	O IQMS foi concebido para ser um framework de mudança do sistema educacional que visa ajudar os professores a identificar suas deficiências e buscar meios de atingir seu desenvolvimento pessoal e profissional. Sofre de falhas de implementação. Fonte: sites do Governo da África do Sul.
Zimbábue	Recursos têm sido direcionados para formação de professores	Possui uma plataforma de coleta de dados sobre professores que facilita uma análise das necessidades dos	Foi introduzido um plano de restauração do status profissional dos professores (EMTP). Fonte: artigos acadêmicos.

	com ênfase na garantia de práticas (foco na avaliação de novos professores enquanto estão se preparando com programa de mentores)	professores, mas não é um método de avaliação per se.	
Namíbia	Cada escola cria um plano estratégico para o desenvolvimento de sua escola. Existe um esquema de certificação de 2 anos para professores iniciantes. Foco é no treinamento de professores enquanto estão estudando.	Combinação de auto-avaliação de professores, com inspeções e surveys de diretores; Uso de novas tecnologias na educação para instrução e avaliação de professores.	Há nesse momento um 'Independent Professional Board' constituído para avaliar a profissionalização dos professores que deve definir parâmetros de registro para professores. Fonte: Banco Mundial.
Zâmbia	Encontra-se no estágio de desenvolvimento de padrões e áreas de prática com apoio da UNESCO.	Não-definidos.	Foram criadas forças-tarefa no governo em 2016 para o desenvolvimento de padrões de monitoramento e avaliação dos professores. Fonte: UNESCO
Botswana	Sistema de avaliação de professores associado a plano de promoções.	Depois de muitas idas e vindas foi introduzido um programa "não-ameaçador" de avaliação de professores no qual os professores aprendem construtivamente a partir de auto-avaliações. Mas foco foi exclusivo em progressão e pagamentos.	Sistema introduzido em 1992 baseado em auto-avaliações foi no final bem recebido pelos professores.
Líbano	Avaliações centralizadas feitas pelo Ministério da Educação a cada dois anos	Avaliações são feitas por entidades externas, como a "Cambridge English Language Assessment", que utilizam surveys com professores, famílias, pais e observações de classe.	Agências da ONU, como a UNESCO e UNICEF tem desenvolvido instrumentos de avaliação dentro de planos mais amplos de reforma educacional. Fonte: UNESCO.
Estados Unidos	Avaliações de professores são conduzidas em vários níveis (sindicatos, governos, ONGs, Fundações)	Originalmente as avaliações eram mais centradas em VA. Recentemente, a política de avaliações tem caminhado na direção de sistemas que incorporam observações de classe e surveys de estudantes (ajudadas por estudos como o MET da Fundação Gates)	O "National Board for Professional Teaching Standards" foi criado em 1987. O seu trabalho é importante não somente para o estabelecimento de rubricas mas para a inclusão de um portfolio de práticas de professores, vídeos, comentários, planos de aula, etc. Esse sistema é utilizado por 32 estados em 500 distritos escolares. Existem muitas inovações locais como os programas "Beginning Educator Support Training" (BEST) de Connecticut ou o "Performance Assessment for California Teachers" (PACT). Fonte: Departamento de Educação dos Estados Unidos.

Canadá	Não existe um departamento nacional de educação no Canadá, mas ONGs como a "Canadian Teachers' Federation" que se encarregam da avaliação de professores.	Avaliações baseadas em observações de classe fazem parte do licenciamento e implementação de padrões profissionais feito pelos 'College of Teachers' nas diferentes regiões. Demanda-se também dos professores a submissão de planos de trabalho discutidos com a direção. Algumas regiões como Ontario têm um sistema de avaliação próprio que envolve reuniões de avaliação, sistemas de notas e relatórios 'refletidos', como são chamados.	Em Alberta foi introduzido um "Teacher Growth, Supervision and Evaluation Policy" onde essas três importantes dimensões foram articuladas em um único conjunto de políticas e ações. Fonte: Governo do Canadá.
Equador	Existem dois padrões de avaliação funcionando na prática: um geral definido pelo Ministério da Educação e outro mais específico pelo Instituto Nacional de Avaliação (INEVAL) dentro do programa 'Ser Maestro'	Combinação de vários instrumentos: auto-avaliação, co-avaliação (um colega escreve sobre o desempenho de outro), testes de práticas online, surveys de diretores sobre professores e surveys de pais e alunos. Tudo é feito na plataforma INEVAL.	A reforma constitucional de 2008 colocou a avaliação de professores no centro das políticas de melhorias educacionais. INEVAL foi criado em 2013. Fonte: sites do Governo do Equador.
México	Há uma Política Nacional de Avaliação de Professores introduzida em 2015 como parte de uma ampla reforma educacional. Está vinculada ao plano de carreira dos professores.	Os parâmetros de avaliação de professores são dados de acordo com o nível de ensino no qual estão os professores. Estão baseados em dois questionários padronizados para professores e direção respondidos em uma plataforma online, seguidos do registro de um projeto de ensino e um exame de conhecimentos didáticos e curriculares.	O sistema de "Evaluación de Desempeño" iniciado em 2015 introduziu um sistema combinado de avaliações com 1) "Informe de Responsabilidades Profesionales", 2)"Proyecto de Enseñaza" e 3) "Examen de conocimientos didáticos y curriculares"
Peru	Existe um modelo de avaliação de desempenho de professores implementado pelo Ministério da Educação.	Combinação de vários instrumentos: observação de classes, visitas a escolas, surveys com famílias e com direções de escolas.	A política de "Marco de Buen Desempeño Docente" foi introduzida em 2012 com um sistema de acreditação de professores e assembleias de governos regionais. Fonte: Ministério de Educação do Peru.
Colômbia	Avaliações centralizadas feitas pelo Ministério da Educação.	Avaliações qualitativas baseadas em auto-avaliações executadas por plataforma online combinadas com gravação de vídeos que são avaliados por outros professores (modelo de observação de classe); Critérios: identificação do contexto sócio-econômico, projeto pedagógico, práxis pedagógica, ambiente de classe.	Programa "Maestro 2025" introduziu uma plataforma virtual onde os professores podem seguir seus processos de avaliação e resultados; Mas avaliações de professores são obrigatórias por lei (art. 84 da Lei Geral de Educação de 1994). Fonte: Ministério da Educação da Colômbia.

Chile	Existe um sistema bem estabelecido para a avaliação de professores baseado no “Marco para la Buena Enseñanza” (MBE) de 2008, que considera quatro domínios (preparo, criação de um ambiente de aprendizado, gestão de classe e responsabilidades profissionais)	Há uma avaliação compulsória (“Sistema de Evaluación del Desempeño Profesional Docente – Docentemás”) e três programas de incentivos (“Asignación Variable por Desempeño Individual” AVDI, “Sistema Nacional de Evaluación del Desempeño” SNED e “Programa de Asignación de Excelencia Pedagógica” AEP). São usados instrumentos variados como auto-avaliação, entrevistas, surveys de professores e diretores, testes de conhecimento dos professores e avaliação de portfólios.	Interessante observar o uso de rankings dentro de ‘grupos homogêneos’ (grupos de escolas comparáveis em termos de características sócio-econômicas de seus estudantes e de características das escolas); O modo pelo qual o MBE coordena essas avaliações pode ser entendido como uma inovação institucional.
Argentina	Não existe um instrumento ou entidade nacional responsável pela avaliação de professores. Avaliações de professores são feitas pelos estados (províncias).	Auto-avaliação online de professores para que reflitam sobre suas práticas. Incluiu um curso online.	Houve em 2016 uma proposta de criação de uma entidade nacional para a avaliação de professores, contudo devido a uma forte oposição dos sindicatos dos professores não saiu do papel. Fonte: Ministério da Educação da Argentina.
Brasil	Não há no Brasil uma política nacional para avaliação de professores. A qualidade dos professores é indiretamente avaliada por indicadores como o IDEB. As Diretrizes Nacionais Curriculares da Educação Básica não fazem referência explícita à avaliação de professores.	Não há um sistema nacional, mas alguns estados como o de São Paulo implementaram (sob a Lei Complementar 1097/2009) um sistema de carreiras que inclui um exame de avaliação de professores. Por outro lado os estados de Bahia e Amazonas estabeleceram sistemas de progressão baseados na participação voluntária de professores em exames de avaliação de desempenho.	O Ministério da Educação instituiu em 2011 a “Prova Nacional de Concurso para o Ingresso na Carreira Docente” (não-compulsória) e a “Matriz de Referência” com competências esperadas de professores qualificados mas essas iniciativas não foram levadas adiante. Fonte: Ministério da Educação do Brasil.
Hong-Kong	Há um instrumento e pilotos para os anos 3 e 6 do ensino fundamental que ainda estão em fase de teste.	O “Bureau” de Educação de Hong Kong (equivalente a secretaria de educação) trabalha com um instrumento chamado “Avaliação do Sistema por Território” (TSA) que avalia o desempenho dos professores indiretamente através das notas dos alunos (VA).	O TSA parece ser uma política não intrusiva mas não representa uma inovação.
Cingapura	Há um sistema de avaliação promovido pelo Ministério da Educação que é parte de um plano de carreira.	Avaliação é feita pelo Ministério da Educação com o instrumento “Enhanced Performance Management System” (EPMS) que dá feedback e coaching aos professores	Introduziu um sistema de ‘tracks’: de ensino, de liderança e de especialista sênior. Fonte: OCDE



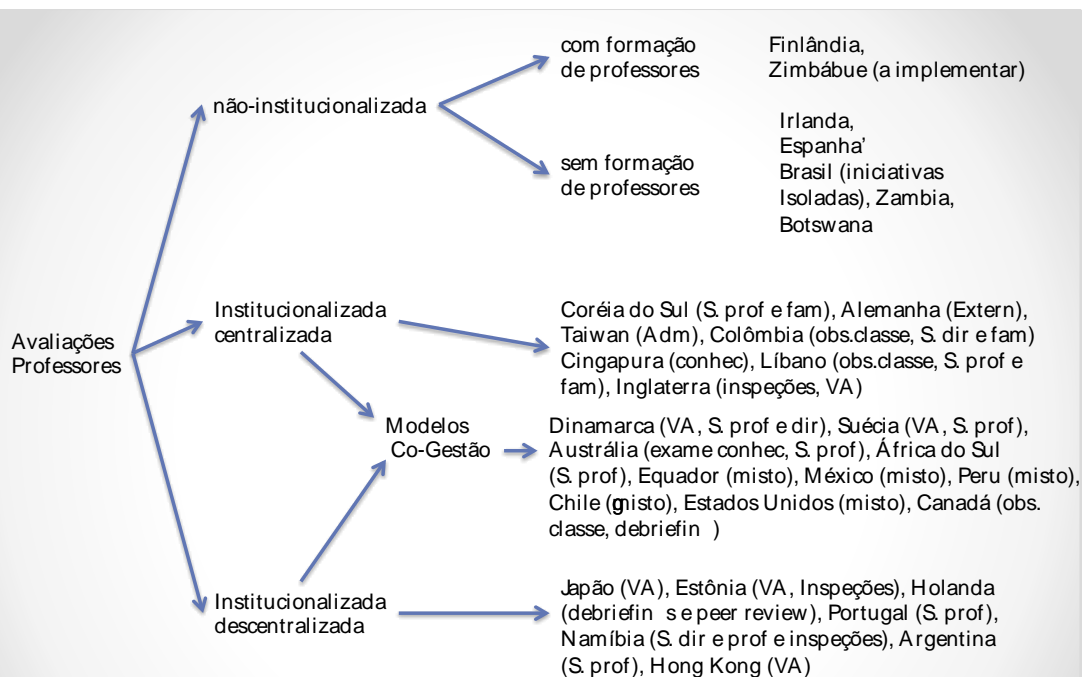
		individualmente. Mede competências dos professores.	
--	--	---	--

As principais características dos sistemas de avaliação descritos acima podem ser sistematizadas segundo duas linhas de critério: i) primeiro, segundo ao tipo de institucionalização dessas medidas (não institucionalizada, institucionalizada centralizada, e institucionalizada descentralizada, possibilitando a existência de modelos de co-gestão) e ii) segundo, de acordo com a existência ou não de programas de formação de professores. Mais do que uma simples classificação de países (que são únicos em suas especificidades históricas e institucionais, veja por exemplo Chen (2015) que mostra como professores chineses são pressionados atualmente a ensinar dentro de novos parâmetros de individualidade e criatividade), o que se busca é uma discussão sobre esses dois critérios como estruturantes das experiências de vários países e definidores de estratégias de formação e avaliação possíveis para países.

Além disso cabe notar que a experiência de diferentes países oferece um laboratório importante no qual podemos refletir sobre alternativas pedagógicas e avaliativas em termos de seus méritos e restrições. Existem contextos institucionais específicos onde valores como 'accountability' se tornam mais relevantes, como relatado por Faubert (2009) no caso da OCDE. Muitos países preferem também focar na avaliação como parte do treinamento de futuros professores (Fehr e Agnello, 2012). O Japão, por exemplo, que há alguns anos atrás fez importantes mudanças nos métodos de avaliação de professores (que usava anteriormente rankings de diretores para classificar professores, passando para observações de classe, auto avaliações e avaliações conjuntas), enfrentou sérios problemas de implementação de novos protocolos. Katsuno (2010) mostrou como deve-se prestar mais atenção para o que denominou 'micropolítica' e 'relações de micro poder' nas escolas para uma implementação bem-sucedida de protocolos de avaliação. Mesmo os países com maior literatura sobre instrumentos de valor-adicionado, como Estados Unidos e Inglaterra, tem debatido muito métodos e instrumentos alternativos de avaliação de seus professores (Looney, 2011).

Claro, idiosincrasias à parte, há lições e elementos importantes em todas essas experiências. O fundamental é sabermos separarmos similaridades e diferenças. Por exemplo, em um estudo feito com professores Chineses e Norte-Americanos (Grant et al, 2013) verificou-se que existem diferenças importantes no modo pelo qual esses professores planejam suas aulas, fazem sua gestão de classe e se relacionam com estudantes e pais, entre outros pontos.

#### **Diagrama 10 – Tipos de avaliações de professores no mundo**



É comum encontrarmos relatos baseados em histórias de implementação de políticas de avaliação de professores em vários países do mundo que não foram muito exitosas devido à falta de aceitação pelos professores. No entanto, é justo dizer que existem também evidências que sugerem a existência de alguns aspectos positivos, como a promoção de debates entre os professores, uma oportunidade de reflexão sobre suas práticas e sobre sua imagem na mídia, sua auto-imagem profissional, e novos desafios que os fizeram repensar o modo como todo o sistema de avaliação de professores está posto, como no caso de Portugal (Flores, 2012).

## 7. Conclusões

Vamos regressar à pergunta fundamental de nosso trabalho. Como avaliar professores? Procuramos responder a essa pergunta examinando os principais instrumentos e métodos utilizados para avaliar professores, especificando suas questões, seus resultados principais, assim como suas limitações. Nesse sentido esse trabalho procurou e procura servir de guia para pessoas interessadas em analisar avaliações feitas de professores e/ou elaborar suas próprias avaliações de professores. Para as pessoas interessadas em analisar as avaliações, esperamos que possam ter encontrado aqui uma discussão sobre outros instrumentos que oferecem outras perspectivas complementares a um determinado tipo de avaliação, assim como questões que um tipo de avaliação particular de interesse deve enfrentar. Para as pessoas interessadas em elaborar suas próprias avaliações, esperamos que possam ter encontrado aqui uma reflexão sobre critérios, protocolos e questões que devem ser percorridos em qualquer tipo de avaliação.

Apresentamos também algumas evidências sobre aspectos concretos das aplicações dos instrumentos de avaliação de professores. Vimos que para o maior estudo já realizado para a comparação de métodos de avaliação de professores (pela Fundação Gates) é melhor usarmos instrumentos mistos para a avaliação de professores ao invés de escolher um único instrumento. Vimos que instrumentos diferentes, não apenas incorporam pressupostos pedagógicos diferentes sobre o que é um professor de qualidade, mas captam dimensões e aspectos distintos do trabalho docente. Mais ainda, vimos que nem todos os métodos de avaliação produzem instrumentos de avaliação profissional para a melhoria da efetividade e crescimento profissional dos professores.

As evidências coletadas sugerem que os sistemas de avaliação de professores no mundo são heterogêneos e que em muitos países os sistemas de avaliação (mesmo os descentralizados) são pouco úteis para a gestão escolar e para o aperfeiçoamento dos professores. De fato, a grande maioria dos países com avaliações centralizadas, usa por conveniência instrumentos de mensuração que avaliam apenas indiretamente a qualidade dos professores (como instrumentos de valor-adicionado) jogando para pesquisadores a responsabilidade de elucidação de problemas de inferência e atribuição do trabalho dos

professores sobre o desempenho dos alunos em testes padronizados (o que exclui algumas matérias importantes como história, artes ou esportes).

A pesquisa também revelou que existem muitos instrumentos além dos tradicionais métodos de valor-agregado, observação de classe e surveys de alunos, como modelos de simulação e protocolos de reflexão e internalização de princípios pedagógicos (como a incorporação de elementos de 'debriefing'). Similarmente, dentro dos instrumentos de observação de classe vemos que as evidências examinadas sugerem modelos alternativos e *guidelines* muito concretos de como estruturar as observações e avaliações para atingir níveis mais altos de intersubjetividade, confiabilidade e validade dos resultados. Vale à pena remarcar esse ponto: os diferentes instrumentos devem ser avaliados segundo os critérios técnicos mais importantes encontrados na literatura, quais sejam, viés, confiabilidade e validade das medidas.

Em relação ao uso potencial dos resultados das avaliações vimos que a lição principal foi de que seu uso depende prioritariamente do tipo de medida utilizada. Instrumentos de percepção ou de observação de classe não têm o mesmo potencial sistemático de medidas de valor-agregado, que por sua vez não tem o potencial de serem usadas para formação de professores. Talvez não existisse tanto objeção a avaliação de professores, como encontramos em alguns países, se seus usos fossem melhor especificados e esclarecidos. Se positivas, as avaliações podem ser usadas para incentivos, promoção e alocação de professores. Se negativas, podem ser usadas para preparar melhor os piores professores ou no limite dispensa-los, o que pode naturalmente causar muita apreensão e incerteza. Por essa razão, objetivos claros *a priori* são fundamentais para a implementação de processos de avaliação de professores. Outra questão não menos importante, como discutida por Liakopoulou (2011) é: quem deve estabelecer esses critérios e instrumentos de avaliação?

Acerca das formações de professores, as evidências coletas sugerem, por exemplo, que sejam baseadas em instruções concretas, com ênfase na gestão de classe, que levem em

conta atributos psicoemocionais (considerados estratégicos nos modelos mais completos de avaliação) e que sejam feitas dentro do ambiente de trabalho. Modelos como o PBTE (do inglês, 'Practice-based Teacher Education, ou seja, Educação dos Professores baseada na prática) podem oferecer bons pontos de partida e inspiração (Ball e Forzani, 2009).

Bons professores e professoras são a chave para uma educação de qualidade. Não avaliá-los não é uma opção. Os problemas educacionais devem ser enfrentados de frente e não escondidos para debaixo do tapete. Mais além disso, avaliações de professores podem abrir novas oportunidades de políticas públicas, como mencionadas acima, baseadas em evidências. Fica aqui, no entanto uma palavra de prudência: para uma visão inclusiva é importante que medidas amplas de avaliação de professores (mistas, se possível) sejam utilizadas porque educar é uma função complexa. Desenhar mecanismos de política pública que ignorem essa complexidade pode levar a distorção de incentivos e efeitos indesejáveis com conseqüente empobrecimento da educação. Esperamos que esse trabalho tenha contribuído para enriquecer o debate sobre instrumentos de avaliação de professores e seu papel para uma educação de melhor qualidade.

## 8. Referências

Abayoumi, A. Arigababu and Oludipe, D.I. (2010) "Perceived Efficacy Beliefs of Nigerian Science Teachers" *Journal of Science Education and Technology*, Vol. 19 No. 1 pp. 27-31

Abu-Tineh, A.M.; Khasawneh, S.A. e Khasawneh H.A. (2011) "Teacher self-efficacy and classroom management styles in Jordanian Schools", *Management Education*, No. 24 Iss. 4 pp. 175-181

Adams, S. e Wolf, K.(2008) "Strengthening the Preparation of Early Childhood Teacher Candidates Through Performance-Based Assessments", *Journal of Early Childhood Teacher Education*, 29:1, 6-29,

Aguilar, C. e Richerme, L. (2014) "What Is Everyone Saying About Teacher Evaluation? Framing the Intended and Inadvertent Causes and Consequences of Race to the Top", *Arts Education Policy Review*, 115:4, 110-120

Alaçam, N. e Olgan, R. (2017) "Pre-service early childhood teachers' self-efficacy beliefs towards parent involvement", *Teaching Education*, 28:4, 421-434

Alford, B.; Rollins, K.; Padro', Y. e Waxman, H. (2016) "Using Systematic Classroom Observation to Explore Student Engagement as a Function of Teachers' Developmentally Appropriate Instructional Practices (DAIP) in Ethnically Diverse Pre-kindergarten Through Second-Grade Classrooms" *Early Childhood Education Journal*, 44:623–635

Amador, J.M. (2017) "Pre-service teachers' video simulations and subsequent noticing: a practice-based method to prepare mathematics teachers." *Research in Mathematics Education*, Vol. 19 pp. 217-235

Amorim-Neto, R.; Rodrigues, V.; Stewart, D.; Xiao, A. e Snyder, J. (2018) “The influence of self-efficacy on entrepreneurial behavior among K-12 teachers”, *Teaching and Teacher Education*, Vol. 72, pp. 44-53

Amrein-Beardsley, A.; Polasky, S. e Holloway-Libell, J. (2016) “Validating ‘value added’ in the primary grades: one district’s attempts to increase fairness and inclusivity in its teacher evaluation system” *Educ Asse Eval Acc*, Vol. 28, pp.139–159

Anagnostopoulos, D.; Levine, T.; Roselle, R. e Lombardi, A. (2018) “Learning to redesign teacher education: a conceptual framework to support program Change”, *Teaching Education*, 29:1, pp.61-80

Anderson-Levitt, K.; van Draanen, J. e Davis, H. (2017) “Coherence, dissonance, and personal style in learning to teach”, *Teaching Education*, 28:4, pp.377-392

Austin, B.; Adesope, O.; French, B.; Gotch, C.; Bélanger, J. e Kubacka, K. (2015), “Examining school context and its influence on teachers: linking Talis 2013 with PISA 2012 student data”, *OECD Education Working Papers*, No. 115,

OECD Publishing, Paris. <http://dx.doi.org/10.1787/5js3f5fgkns4-en>

Backer, E. (2012) “Burnt at the Student Evaluation Stake – the penalty for failing students”, *e-Journal of Business Education & Scholarship of Teaching* Vol. 6, No. 1, 2012, pp: 1- 13.



Baecher, L. e Kung, S. (2011) “Jumpstarting Novice Teachers’ Ability to Analyze Classroom Video”, *Journal of Digital Learning in Teacher Education*, 28:1, pp.16-26

Bailey, L. (2010) “The Impact of Sustained, Standards-Based Professional Learning on Second and Third Grade Teachers’ Content and Pedagogical Knowledge in Integrated Mathematics”, *Early Childhood Educ J*, Vol. 38, pp.123–132

Ball, D. e Forzani, F. (2009). “The work of teaching and the challenge for teacher Education”, *Journal of Teacher Education*, Vol. 60, pp. 497–511.

Ballou, D. (2009) “Test Scaling and Value-Added Measurement”, *Education Finance and Policy*, Vol. 4, No. 4, Key Issues in Value-Added Modelling, pp. 351-383

Balossi, M. e Hernandez, N. (2016) “On Teacher Quality in Independent Schools”, *Peabody Journal of Education*, Vol.91, n.5, pp.672-690

Banister, S. e Reinhart, R. (2012) “Assessing NETS • T Performance in Teacher Candidates”, *Journal of Digital Learning in Teacher Education*, Vol. 29, n.2, pp.59-65

Barrera-Pedemonte, F. (2016), “High-Quality Teacher Professional Development and Classroom Teaching Practices: Evidence from Talis 2013”, *OECD Education*

*Working Papers*, No. 141, OECD Publishing, Paris. <http://dx.doi.org/10.1787/5jlpszw26rvd-en>

Bélaire, L. e Dionne, E. (2009) “Evaluation des apprentissages dans le contexte

québécois : Entre décisions politiques et pratiques en salle de classe”, *Mesure et évaluation en éducation*, Vol. 32, n. 3, pp. 76-100

Belsito, C. (2016) “The importance of teacher quality and quality teaching on academic performance” *Journal of Student Engagement Education Matters*, Vol. 6 pp. 26-38

Bierman, K.; DeRousie, R.; Heinrichs, B.; Domitrovich, C.; Greenberg, M. e Gill, S. (2013) “Sustaining High-Quality Teaching and Evidence-Based Curricula: Follow-Up Assessment of Teachers in the REDI Project”, *Early Education & Development*, Vol. 24, n.8, pp.1194-1213

Bonner, S.; Rivera, C. e Chen, P. (2018) “Standards and assessment: coherence from the teacher’s perspective” *Educ Asse Eval Acc*, Vol. 30, pp.71–92

Borrero, N. e Sanchez, G. (2017) “Enacting culturally relevant pedagogy: asset mapping in urban classrooms”, *Teaching Education*, Vol. 28, n.3, pp. 279-295

Bowen, B. (2013) “Measuring Teacher Effectiveness When Comparing Alternatively and Traditionally Licensed High School Technology Education Teachers in North Carolina”, *Journal of Technology Education*, Vol. 25, n.1, pp. 82-100

Bozkus, K. and Tastan, M. (2016) “Teacher Opinions about Qualities of Effective Teaching.” *Turkey Faculty of Education*, Vol.6 No. 4 pp. 469-490

Briggs, D. e Dadey, N. (2017) “Principal holistic judgments and high-stakes

evaluations of teachers”, *Educ Asse Eval Acc*, Vol. 29, pp. 155–178

Brocato, B.; Bonanno, A. e Ulbig, S. (2015) “Student perceptions and instructional evaluations: A multivariate analysis of online and face-to-face classroom settings”, *Educ Inf Technol*, Vol. 20, pp. 37–55

Bush, B.; Frank, I. e Dixon-Krauss, L. (2014) “Assessing Beyond Minimal Compliance”, *Action in Teacher Education*, Vol.36, n. 5-6, pp. 351-362

Buttner, S.; Pijl, S.; Bijstra, J. e van den Bosch, E. (2015) “Triangulating Measures of Teacher Quality in Teaching Students with Behavioral Problems”, *Journal of Cognitive Education and Psychology*, Vol. 14, n. 3, pp. 294-313

Byers, A.; Koba, S.; Sherman, G.; Scheppke, J. e Bolus, R. (2011) “Developing a Web-Based Mechanism for Assessing Teacher Science Content Knowledge”, *J Sci Teacher Educ*, Vol. 22, pp. 273–289

Camburn, E.M. (2010) “Embedded Teacher Learning Opportunities as a Site for Reflective Practice: An Exploratory Study.” *American Journal of Education*, Vol. 116 No. 4 pp. 463-489

Cantrell, S. (2012) “The Measures of Effective Teaching Project: An Experiment to Build Evidence and Trust”, *Education Finance and Policy*, Vol. 7, No. 2, pp. 203-218

Carter, L.; Wiliford, A. e LoCasale-Crouch, J. (2014) “Reliability and Validity of a Measure of Preschool Teachers’ Attributions for Disruptive Behavior”, *Early Education and Development*, Vol. 25, pp. 949–972

Cassidy, D.; Lower, J.; Kintner, V. e Hestenes, L. (2009) "Teacher Ethnicity and Variation in Context: The Implications for Classroom Quality", *Early Education and Development*, vol. 20, n. 2, pp.305-320

Caughlan, S. e Jiang, H. (2014) "Observation and Teacher Quality: Critical Analysis of Observational Instruments in Preservice Teacher Performance Assessment", *Journal of Teacher Education*, Vol. 65, n. 3, pp. 375-388.

Champagne, M. (2013) "Student use of mobile devices in course evaluation: a longitudinal study", *Educational Research and Evaluation*, Vol.19, n.7, pp.636-646

Chan, C.; Luk, L. e Zeng, M. (2014) "Teachers' perceptions of student evaluations of teaching", *Educational Research and Evaluation*, Vol.20, n.4, pp. 275-289

Chandrasekhar, A.; Durazo-Arvizu, R.; Hoyt, A. e McNulty, J. (2013) "Do student evaluations influence the teaching skills of clerkship clinical faculty?", *Educational Research and Evaluation*, Vol. 19, n.7, pp. 628-635

Charalambous, Y. (2016) "Investigating the Knowledge Needed for Teaching Mathematics: An Exploratory Validation Study Focusing on Teaching Practices." *Journal of Teaching Education*, 67;3, pp. 220-237

Charteris, J. e Smardon, D. (2015) "Teacher agency and dialogic feedback: Using classroom data for practitioner inquiry", *Teaching and Teacher Education*, Vol. 50, pp. 114-123

Chen, J. (2015) "Teachers' Conceptions of Approaches to Teaching: A Chinese Perspective",

Chen, W.; Hendricks, K. e Archibald, K. (2011) "Assessing pre-service teachers' quality teaching practices", *Educational Research and Evaluation*, Vol.17, n.1, pp.13-32

Chen, W.; Mason, S.; Staniszewski, C., Upton, A. e Valley, M. (2012) "Assessing the quality of teachers' teaching practices", *Educ Asse Eval Acc*, Vol. 24, pp. 25–41

Cheong, G. (2017) "Developing an Instrument for Self-Evaluation of Teaching and Learning Competencies: A Review of Faculty Professional Development and the Changing Higher Education Landscape in Singapore", *Education Research and Perspectives*, Vol. 44, pp. 1-20

Cherrington, S. e Loveridge, J. (2014) "Using video to promote early childhood teachers' thinking and reflection", *Teaching and Teacher Education*, Vol. 41, pp. 42-51

Coles, A. and Scott, H. (2015) "Planning for the unexpected in the mathematics classroom: and account of teacher and student change." *Research in Mathematics Education*, Vol. 17, n.2, pp. 128-147

Curby, T.W. Boyer, C. Edwards, T. and Chavez, C. (2012) "Assistant Teachers in Head Start Classrooms: Comparing to and Working with Lead Teachers." *Early Education and Development*, 23;5 pp. 640-653

Dalioglu, S. e Adiguzel, O. (2016) "Teacher candidates' self-efficacy beliefs and possible selves throughout the teaching practice period in Turkey", *Asia Pacific Educ. Rev.*, Vol. 17, pp. 651–661

Darling-Hammond, L.; Amrein-Beardsley, A.; Haertel, E. e Rothstein, J. (2012). "Evaluating teacher evaluation". *Phi Delta Kappan*, Vol. 93, n. 6, pp. 8–15

David, A. e Zoch, M. (2015) "Understanding Teachers' Perspectives on Being Researched: A Case Study of Two Writing Teachers", *The Journal of Writing Teaching Education*, Vol. 4, issue 2, pp. 161-181

Davidovitch, N.; Soen, D. e Sinuani-Stern, Z. (2011) "Performance measures of academic faculty – a case study", *Journal of Further and Higher Education*, Vol. 35, n.3, pp.355-373

DeGraff, T.; Schmidt, C. e Waddell, J. (2015) "Field-based teacher education in literacy: preparing teachers in real classroom contexts", *Teaching Education*, Vol. 26, n. 4, pp. 366-382

Delport, A. e Makaye, J. (2009) "Clustering schools to improve teacher professional development: Lessons learnt from a Zimbabwean case study", *Africa Education Review*, Vol.6, n.1, pp.96-105

DeLuca, C. (2012) "Preparing Teachers for the Age of Accountability: Toward a Framework for Assessment Education", *Action in Teacher Education*, Vol. 34, n. 5-6, pp. 576-591

Denny, J.; Hallam, R. e Homer, K. (2012) "A Multi-Instrument Examination of Preschool Classroom Quality and the Relationship Between Program, Classroom, and Teacher Characteristics", *Early Education & Development*, Vol. 23, n. 5, pp. 678-696

Detroz, P. e Blais, J.G. (2012) "Les étudiants sont-ils compétents pour évaluer la qualité des enseignements ?" *Mesure et évaluation en éducation*, Vol. 35, N. 3, pp. 5-30

De Witte, K. e Van Klaveren, C. (2014) "How are teachers teaching? A nonparametric approach", *Education Economics*, Vol. 22, n.1, pp. 3-23

Dixon, M. (2016) "Teacher beliefs in 'testing' times: a lesson from Singapore", *Teaching Education*, Vol. 27, n. 3, pp. 327-339

Duquette, C. e Dabrowski, L. (2016) "A Study of classroom inquiry and reflection among preservice teachers candidates", *McGill Journal of Education*, Vol. 51, n. 1, pp. 575-595

Echávarri, J., & Peraza, C. (2017). "Modernizing schools in Mexico: The rise of teacher assessment and school-based management policies", *Education Policy Analysis Archives*, Vol. 25, n.90, pp. 1-25

Echazarra, A. et al. (2016), "How teachers teach and students learn: Successful strategies for school", *OECD Education Working Papers*, No. 130, OECD Publishing, Paris.

<http://dx.doi.org/10.1787/5jm29kpt0xxx-en>

Ekici, D. (2018) "Development of pre-service teachers' teaching self-efficacy beliefs through an online community of practice", *Asia Pacific Education Review*, Vol. 19, pp. 27–40

Elstad, E.; Lejonberg, E. e Christophersen, K. (2015) "Teaching evaluation as a contested practice: Teacher resistance to teaching evaluation schemes in Norway", *Education Inquiry*, Vol. 6, n. 4, pp. 375-399

Erdemir, N. (2011) "Primary school teachers' levels of self-efficacy according to various settlement areas", *Asia Pacific Educ. Rev.*, Vol. 12, pp. 237–249

Farley, K.; Piasta, S.; Dogucu, M. e O'Connell, A. (2017) "Assessing and Predicting Small-Group Literacy Instruction in Early Childhood Classrooms", *Early Education and Development*, Vol. 28, n. 4, pp. 488-505

Faubert, V. (2009), "School Evaluation: Current Practices in OECD Countries and a Literature Review", *OECD Education Working Papers*, No. 42, OECD publishing.

Fehr, M. e Agnello, M. (2012) "Engaging in Diverse Classrooms Using a Diversity Awareness Survey to Measure Preservice Teachers' Preparedness, Willingness, & Comfort", *Multicultural Education Research*, pp. 34-39



Fernández, E.; LeChasseur, K. e & Donaldson, M. (2018) “Responses to including parents in teacher evaluation policy: A critical policy analysis”, *Journal of*

*Education Policy*, Vol. 33, n. 3, pp. 398-413

Ferreira, L. (2017) “Desenvolvimento Profissional e carreira docente: diálogos sobre professores iniciantes”, *Acta Scientiarum Education*, Vol. 39, n. 1, pp. 79-89

Flores, M. (2012) “The implementation of a new policy on teacher appraisal in Portugal: how do teachers experience it at school?” *Educ Assessment Eval Acc*, Vol. 24, pp. 351-368

Ford, T. G., Urick, A., e Wilson, A. S. P. (2018). “Exploring the effect of supportive teacher evaluation experiences on U.S. teachers’ job satisfaction”, *Education Policy Analysis Archives*, Vol. 26, n. 59, pp. 1-36

Gargani, J. e Strong, M. (2014) “Can We Identify a Successful Teacher Better, Faster, and Cheaper? Evidence for Innovating Teacher Observation Systems”, *Journal of Teacher Education*, Vol. 65, n. 5, pp. 389-401

Gaudreau, N., Frenette, É. e Thibodeau, S. (2015). “Élaboration de l’Échelle du sentiment d’efficacité personnelle des enseignants en gestion de classe (ÉSEPGC)”, *Mesure et évaluation en éducation*, Vol. 38, n. 2, pp. 31–60

Giallousi, M.; Tselfes, V. e Gialamas, V. (2014) “Using Student- Teachers’ Reports of Self-Efficacy to Evaluate an Early Childhood Science Course”, *Journal of Early*

*Childhood Teacher Education*, Vol. 35, n. 4, pp. 337-356

Gibson, D. (2013) "Assessing Teaching Skills with a Mobile Situation." *Journal of Digital Learning in Teacher Education*, Vol. 30, n. 1, pp. 4-10

Goddard, R.; Hoy, W. e Hoy, A. (2000) "Collective Teacher Efficacy: its meaning, measure, and impact on student achievement", *American Educational Research Journal*, Vol. 37, n. 2, pp. 479-507

Goodman, G.; Arbona, C. e Rameriz, R. (2008) "High-Stakes, Minimum-Competency Exams: how competent are they for evaluating teacher competence?" *Journal of Teacher Education*, Vol. 59, n. 1, pp. 24-39

Gossman, P e Horder, S. (2016) "Effective teacher? Student self- evaluation of development and progress on a teacher education programme", *Journal of Further and Higher Education*, Vol. 40, n. 4, pp. 447-465

Grant, L.; Stronge, J. e Xu, X. (2013) "A Cross-cultural comparative study of teacher effectiveness: analyses of award-winning teachers in the United States of America", *Educ Asse Eval Acc*, Vol. 25, pp. 251-276

Green, C.; Medina-Jerez, W. e Bryant, C. (2016) "Cultivating environmental citizenship in teacher education", *Teaching Education*, Vol. 27, n. 2, pp. 117-135

Grossman, P., Loeb, S., Cohen, J., & Wyckoff, J. (2013). Measure for measure: The relationship between measures of instructional practice in middle school English language arts and teachers' value-added scores. *American Journal of Education*. 119(3), 445-470

Grossman, Pam, Sharon Greenberg, Karen Hammerness, Julie Cohen, Chandra Alston, and Michelle Brown. (2009). "Development of the Protocol for Language Arts Teaching Observation (PLATO)." Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA, April

Ha, B. e Sung, Y. (2011) "Teacher reactions to the performance-based bonus program: how the expectancy theory Works in the South Korean School culture", *Asia Pacific Educ Review*, Vol. 12, pp. 129-141

Haim, O. e Amdur, L. (2016) "Teacher perspectives on their alternative fast-track induction", *Teaching Education*, Vol. 27, n. 4, pp. 343-370

Halim, L.; Abdullah, S. e Meerah, T. (2014) "Students' Perceptions of Their Science Teachers' Pedagogical Content Knowledge", *Journal of Science Education and Technology*, Vol. 23, No. 2, pp. 227-237

Hallinger, P. (2010) "Using Faculty Evaluation to improve teaching quality: a longitudinal case study of higher education in Southeast Asia", *Educ Asse Eval Acc*, Vol. 22, pp. 253-274

Hallinger, P.; Heck, R. e Murphy (2014) "Teacher evaluation and school improvement: an analysis of the evidence", *Educ Asse Eval Acc*, Vol. 26, pp. 5-28

Hamdan, A.D. Anuar, M.K. and Khan, A. (2016) "Implementation of co-teaching approach in an inclusive classroom: overview of the challenges, readiness, and role of special education teacher." *Asia Pacific Educ. Rev*, Vol. 17 pp. 289-298

Hamilton, L.; Stecher, B. e Yuan, K. (2012) "Standards-Based Accountability in the United States:", *Education Inquiry*, Vol. 3, n. 2, 149-170

Han, S. (2016) "Teaching diversity: a reflexive learning opportunity for a teacher education", *Teaching Education*, Vol. 27, n. 4, pp. 410-426

Hanushek, Eric, and Steven Rivkin. 2010. Using Value-Added Measures of Teacher Quality. Brief No. 9. Washington, DC: CALDER.

Harris, D. (2009) "Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives", *Education Finance and Policy*, Vol. 4, n. 4, pp. 319-350

Hawley, L.; Bovaird, J. e Wu, C. (2017) "Stability of Teacher Value-Added Rankings Across Measurement Model and Scaling Conditions", *Applied Measurement in Education*, Vol. 30, n. 3, pp. 196-212

Hayes, K.; Lee, C.; SiStefano, R.; O'Connor, D. e Seitz, J. (2016) "Measuring Science Instructional Practice: a survey tool for the age of NGSS", *J Sci Teacher Education*, Vol. 27, pp. 137-164

Henderson, C. Turpen, C. Dancy, M. and Chapman, T. (2014) "Assessment of Teaching effectiveness: Lack of alignment between instructors, institutions, and research recommendations", *Physics Education Research*, Vol. 10, pp. 1-20

Bjerke, A. e Eriksen, E. (2016) "Measuring pre-service teachers' self-efficacy in tutoring children in primary mathematics: an instrument", *Research in Mathematics Education*, Vol. 18, n. 1, pp. 61-79

Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794-831.

Hill, H. C., Umland, K., Litke, E., & Kapitula, L. R. (2012). Teacher quality and quality teaching: Examining the relationship of a teacher assessment to practice. *American Journal of Education*, 118(4), 489-519.

Hill, H. C., Blunk, M., Charalambous, C., Lewis, J., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical Knowledge for Teaching and the Mathematical Quality of Instruction: An exploratory study. *Cognition and Instruction*, 26, 430-511.

Hill, H. C., Charalambous, C. Y., Blazar, D., McGinn, D., Kraft, M. A., Beisiegel, M., Humez, A., Litke, E., & Lynch, K. (2012). Validating arguments for observational instruments: Attending to multiple sources of variation. *Educational Assessment*, 17(2-3), 88-106.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: teacher observation systems and a case for the generalizability study. *Educational Researcher*, 41(2), 56-64.

Hindin, A. e Mueller, M. (2016) "Creating home-school partnerships: examining urban and suburban teachers' practices, challenges, and educational needs", *Teaching Education*, Vol. 27, n. 4, pp. 427-445

Hindman, A.; Skibbe, L. e Morrison, F. (2013) “Teacher Outreach to Families across the transition to school: an examination of teachers’ practices and their unique contributions to children’s early academic outcomes”, *Early Childhood Education Journal*, Vol. 41, pp. 391-399

Hobson, A.; Ashby, P.; McIntyre, J. e Malderez, A. (2010), “International Approaches to Teacher Selection and Recruitment”, *OECD Education*

*Working Papers*, No. 47, OECD Publishing.

Hoque, K.; Alam, G. e Abdullah, A. (2011) “Impact of teachers’ professional development on school improvement – an analysis at Bangladesh standpoint”, *Asia Pacific Educ Rev*, Vol. 12, pp. 337-348

Hoyt, K. and Terrantino, J. (2015) “Rethinking Field Observations: Strengthening Teacher Education Through INFORM”, *Action in Teacher Education*, Vol. 37, n. 3, pp. 209-222

Hu, B.; Fan, X.; Gu, C. e Yang, N. (2016) “Applicability of the Classroom Assessment Scoring System in Chinese Preschools Based on Psychometric Evidence”, *Early Education and Development*, Vol. 27, n. 5, pp. 714-734

Hu, B.; Zhou, Y. e Li, K. (2014) “Pinpointing Chinese Early Childhood Teachers’ Professional Development Needs Through Self-Evaluation and External

Observation of Classroom Quality”, *Journal of Early Childhood Teacher Education*, Vol. 35, n.1, pp. 54-78

Huber, S. e Skedsmo, G. (2016) "Teacher Evaluation – accountability and improving teaching practices", *Edu Asse Eval Acc*, Vol. 28, pp. 105-109

Hui, S.; Brown, G. e Chan, S. (2017) "Assessment for learning and for accountability in classrooms: the experience of four Hong Kong primary school curriculum leaders", *Asia Pacific Edu Review*, Vol. 18, pp. 41-51

Humphries, C.; Hebert, E.; Daigle, K. e Martin, J. (2012) "Development of a Physical Education Teaching Efficacy Scale", *Measurement in Physical Education and Exercise Science*, Vol. 16, n. 4, pp. 284-299

Instefjord, E. e Munthe, E. (2017) "Educating digitally competent teachers: A study of integration of professional digital competence in teacher education", *Teaching and Teacher Education*, Vol. 67, pp. 37-45

Ishii, J. e Rivkin, S. (2009) "Impediments to the Estimation of Teacher Value Added", *Education Finance and Policy*, Vol. 4, n. 4, pp. 520-536

Jaber, L.; Southerland, S. e Dake, F. (2018) "Cultivating epistemic empathy in preservice teacher education", *Teaching and Teacher Education*, Vol. 72, pp. 13-23

Jacobs, F.; Knoppers, A. e Webb, L. (2013) "Making sense of teaching social and moral skills in physical education", *Physical Education and Sport Pedagogy*,

Vol. 18, n. 1, pp. 1-14

Jarvis, P. (2011) "What Makes a Good teacher?", *Lifelong Education and Libraries*, Vol. 11, pp. 147-152 (Kyoto University)

Jung, H. e Choi, E. (2016) "The importance of indirect teaching behaviour and its educational effects in physical education", *Physical Education and Sport Pedagogy*, Vol. 21, n. 2, pp. 121-136

Kelcey, B., McGinn, D., & Hill, H. (2014). Approximate measurement invariance in cross-classified rater-mediated assessments. *Frontiers in psychology*, 5, 1469

Kane, T. (2012) "Capturing the dimensions of effective teaching", *J. Education Next*, Vol. 12, issue 4, pp. 1-10

Kangas, M.; Siklander, P.; Randolph, J. e Ruokamo, H. (2017) "Teachers' engagement and students' satisfaction with a playful learning environment", *Teaching and Teacher Education*, Vol. 63, pp. 274-284

Katsuno, M. (2010) "Teacher evaluation in Japanese schools: an examination from a micro - political or relational viewpoint", *Journal of Education Policy*, Vol. 25, n. 3, pp. 293-307

Kazempour, M. e Sadler, T. (2015) "Pre-service teachers' science beliefs, attitudes, and self-efficacy: a multi-case study", *Teaching Education*, Vol. 26, n. 3, pp. 247-271



Kelting-Gibson, L. Gallavan, N.P. ST. Arnauld, E. Black, G. Cayson, A. Davis, J. Evans, K.D. Johnson, P.P. Levandowski, B. Mosley, K. Rickey, D. Shulsky, D.D. Thomas, D. Williamson, A.M. and Wolfgang, J.I. (2014) "Four Facets of Classroom Assessments: Obstacles, Obligations, Outcomes and Opportunities" *Action in Teacher Education*, Vol. 36 No. 5-6

Kerr, S.; Lucas, L.; DiDomenico, G.; Mishra, V.; Stanton, B.; Shivde, G.; Pero, A.; Runyen, M. e Terry, G. (2017) "Is mindfulness training useful for pre-service teachers? An exploratory investigation", *Teaching Education*, Vol. 28, n. 4, pp. 349-359

Kersting, N.; Chen, M. e Stigler, J. (2013) "Value-added teacher estimates as part of teacher evaluations: exploring the effects of data and model specifications on the stability of teacher value-added scores", *Education Policy Analysis Archives*, Vol. 21, n. 7, pp. 1-33

Kilday, C. e Kinzie, M. (2009) "An Analysis of Instruments that measure the quality of mathematics teaching in early childhood", *Early Childhood Educ Journal*, Vol. 36, pp. 365-372

Kim, R.; Sihm, H. e Mitchell, R. (2014) "South Korean Elementary Teachers' mathematics teaching efficacy beliefs: implications for educational policy and research", *Mathematics Education Trends and Research*, pp. 1-17

Kirchner, J.; Evans, C. e Norman, A. (2010) "Examining the Relationship between Two Assessments of Teacher Effectiveness", *Action in Teacher Education*, Vol. 32, n. 1, pp. 73-81

Kitchen, H., et al. (2017), "Romania 2017", *OECD Reviews of Evaluation and Assessment in Education*, OECD Publishing, Paris.

Klassen, R.; Durksen, T.; Hashmi, W.; Kim, L.; Longden, K.; Metsapelto, R.; Poikkeus, A. e Gyori, J. (2018) "National context and teacher characteristics: exploring the critical non-cognitive attributes of novice teachers in four countries", *Teaching and Teacher Education*, Vol. 72, pp. 64-74

Knezek, G.; Hopper, S.; Christensen, R.; Tyler-

Wood, T. e Gibson, D. (2015) "Assessing Pedagogical Balance in a Simulated Classroom Environment", *Journal of Digital Learning in Teacher Education*, Vol. 31, n. 4, pp. 148-159

Koedel, C. e Betts, J. (2011) "Does Student Sorting Invalidate Value-added models of teacher effectiveness? An extended analysis of the Rothstein critique", *Education Finance and Policy*, Vol. 6, n. 1, pp. 18-42

Koedel, C. e Betts, J. (2010) "Value-added to what? How a ceiling in the testing instrument influences value-added estimation", *Education Finance and Policy*, Vol. 5, pp. 54-81

Koirala, H.; Davis, M. e Johnson, P. (2008) "Development of a performance assessment task and rubric to measure prospective secondary school mathematics teachers' pedagogical content knowledge and skills", *J Math Teacher Edu*, Vol. 11, pp. 127-138

Kordts-Freudinger, R. e Geithner, E. (2013) "When mode does not matter: evaluation in class versus out of class", *Educational Research and Evaluation*, Vol. 19, n. 7, pp. 605-61

Kose, A., Uzun, M. (2018). "Prospective Teachers' Views on Effective Teacher

Characteristics and Their Occupational Self-efficacy Perceptions in terms of These Characteristics”, *Educational Process: International Journal*, 7(1), 60-73.

Kyrgiridis, P.; Derri, V.; Emmanouilidou, K.; Chlapoutaki, E. e Kioumourtzoglou, E. (2014) “Development of a Questionnaire for Self- Evaluation of Teacher Effectiveness in Physical Education (SETEQ-PE)”, *Measurement in Physical Education and Exercise Science*, Vol. 18, n. 2, pp. 73-90

La Paro, K.M. Williamson, A.C. and Hatfield, B. (2014) “Assessing Quality in Toddler Classrooms Using the CLASS-Toddler and the ITERS-R”, *Early Education and Development*, No. 25, pp. 875-893

La Paro, K.; Maynard, C.; Thomason, A. e Scott- Little, C. (2012) “Developing Teachers' Classroom Interactions: A Description of a Video Review Process for Early Childhood Education Students”, *Journal of Early Childhood Teacher Education*, Vol. 33, n. 3, pp. 224-238

LaBoskey, V. e Richert, A. (2015) “Self-Study as a Means for Urban Teachers to Transform Academics”, *Studying Teacher Education*, Vol. 11, n. 2, pp. 164-179

Lai, E. (2014) “Principal leadership practices in exploiting situated possibilities to build teacher capacity for change”, *Asia Pacific Educ. Review*, Vol. 15, pp. 165-175

Larose, F.; Grenon, V.; Bédard, J. e Bourque, J. (2009). “Analyse des pratiques enseignantes et la construction d'un référentiel de compétences : perspectives et contraintes méthodologiques”, *Nouveaux cahiers de la recherche en éducation*, Vol. 12, n.1, pp. 65–82

Laveault, D. (2009). “L'évaluation en classe : Des politiques aux pratiques”, *Mesure et évaluation en éducation*, Vol. 32, n. 3, pp. 1–22

Le Donné, N.; Fraser, P. e Bousquet, G. (2016), "Teaching Strategies for Instructional Quality: Insights from the TALIS-PISA Link Data", *OECD Education Working Papers*, No. 148, OECD Publishing, Paris.

Lee, C.; Walkowiak, T. e Nietfeld, J. (2017) "Characterization of mathematics instructional practises for prospective elementary teachers with varying levels of self-efficacy in classroom management and mathematics teaching", *Math Education Research*, Vol. 29, pp. 45-72

Liakopoulou, M. (2011) "Teachers' Pedagogical competence as a prerequisite for entering the profession", *European Journal of Education*, Vol. 46, n. 4, pp. 474-488

Liberante, L. (2012) "The importance of teacher–student relationships, as explored through the lens of the NSW Quality Teaching Model", *Journal of Student Engagement: Education Matters*, Vol. 2, n. 1, pp. 2-9

Lillvist, A.; Sandberg, A.; Sheridan, S. e Williams, P. (2014) "Preschool teacher competence viewed from the perspective of students in early childhood teacher education", *Journal of Education for Teaching*, Vol. 40, n. 1, pp. 3-19

Liu, S.; Xu, X. e Stronge, J. (2016) "Chinese middle school teachers' preferences regarding performance evaluation measures", *Educ Asse Eval Acc*, Vol. 28, pp. 161-177

Liu, S. e Zhao, D. (2013) "Teacher evaluation in China: latest trends and future directions", *Educ Asse Eval Acc*, Vol. 25, pp. 231-250

Looney, J. (2011) "Developing high-quality teachers: teacher evaluation for improvement", *European Journal of Education*, Vol. 46, n. 4, pp. 440-455

Love, T.; Wells, J. e Parkes, K. (2017) "Examining the teaching of science, and technology and engineering content and practices: an instrument modification study", *Journal of Technology Education*, Vol. 29, n. 1, pp. 45-65

Mangiante, E. (2011) "Teachers matter: measures of teacher effectiveness in low-income minority schools", *Educ Asse Eval Acc*, Vol. 23, pp. 41-63

Manier, C. & Abernot, Y. (2016). "Contribution à l'évaluation de l'activité professionnelle sur la base d'une CAH : le cas des éducateurs spécialisés", *Mesure et évaluation en éducation*, Vol. 39, n. 2, pp. 39–58

Martell, C. (2016) "Teaching emerging teacher-researchers: examining a district-based professional development course", *Teaching Education*, Vol. 27, n. 1, pp. 88-102

Maulana, R. e Opdenakker, M. (2014) "Teachers' interpersonal involvement as a predictor of students' academic motivation among Indonesian secondary school students: a multilevel growth curve analysis", *Asia-Pacific Edu Res*, Vol. 23, n. 3, pp. 591-603

McCaffrey, D.; Sass, T.; Lockwood, J. e Mihaly K. (2009) "The intertemporal variability of teacher effect estimates", *Education Finance and Policy*, Vol. 4, n. 4, pp. 572-606

McConnell, T.; Parker, J. e Eberhardt, J. (2013) "Assessing Teachers' science content knowledge: a strategy for assessing depth of understanding", *J Sci Teacher Education*, Vol. 24, pp. 717-743

McCurdy, K., Reagan, E. M., Schram, T., & Rogers, A. (2018). "Integrating performance assessments across a PK-20 continuum: A locally developed collaboration", *Education Policy Analysis Archives*, Vol. 26, n. 14, pp. 1-5

McGuire, P.; Kinzie, M.; Thunder, K. e Berry, R. (2016) "Methods of analysis and overall mathematics teaching quality in at-risk prekindergarten classrooms", *Early Education and Development*, Vol. 27. N. 1, pp. 89-109

McInnes, B. (2017) "Preparing teachers as allies in Indigenous education: benefits of an American Indian content and pedagogy course", *Teaching Education*, Vol. 28, n. 2, pp. 145-161

Meierdirk, C. (2018) "The impact of the social environment on the student teacher's agency", *Teaching Education*, Vol. 29, n. 1, pp. 33-46

Miao, Z.; Reynolds, D.; Harris, A. e Jones, M. (2015) "Comparing performance: a cross-national investigation into the teaching of mathematics in primary classrooms in England and China, Asia Pacific", *Journal of Education*, Vol. 35, n.3, pp. 392-403

Milner, A.R. Sondergeld, T.A Demir, A Johnson, C.C. and Czerniak, C.M. (2011/2012) "Elementary Teacher's Beliefs About Teaching Science and Classroom Practice: An Examination of Pre/Post NCLB Testing in Science", *J Sci Teacher Educ.*, Vol. 23, pp. 111-132

Mincu, M.E. (2015) "Teacher quality and school improvement: what is the role of research?", *Oxford Review of Education*, 41;2, pp. 253-69

Modiba, M. e Nsibande, R. (2008) "Continuous assessment as a selfinitiated teacher development strategy in Swaziland: practice and challenges for effective implementation", *Africa Education Review*, Vol. 5, n. 1, pp. 68-83

Molina, O. e Ratté, S. (2017) "Students and teacher academic evaluation perceptions: methodology to construct a representation based on actionable knowledge discovery framework", *Educ Inf Technol*, Vol. 22, pp. 1043-1066

Morgado, J. e Sousa, F. (2010) "Teacher evaluation, curricular autonomy and professional development: trends and tensions in the Portuguese educational policy", *Journal of Education Policy*, Vol. 25, n. 3, pp. 369-384

Morrisette, J. & Nadeau, M. (2011). "Une typologie du savoir-faire enseignant en matière d'évaluation formative", *Mesure et évaluation en éducation*, Vol. 34, n. 3, pp. 5–25

Moyer-Packenham, P.S. Bolyard, J.J. Kitsantas, A. and Hana Oh (2008) "The Assessment of Mathematics and Science Teacher Quality", *Peabody Journal of Education*, Vol. 83, No. 4, pp. 562-591

Murphy, C. e Martin, S. (2015) "Coteaching in teacher education: research and practice", *Asia-Pacific Journal of Teacher Education*, Vol. 43, n. 4, pp. 277-280

Myers, D.; Sugai, G.; Simonsen, B. e Freeman, J. (2017) "Assessing Teachers' behaviour support skills", *Teacher Education and Special Education*, Vol. 40, n. 2, pp. 128-139

Nagro, S. e Cornelius, K. (2013) "Evaluating the evidence base of video analysis: a special education teacher development tool", *Teacher Education and Special Education*, Vol. 36, n. 4, pp. 312-329

Neel, M. (2017) "Making sense and facing tensions: an investigation of core practice complexities", *Teaching Education*, Vol. 28, n. 3, pp. 257-278

Newton, X.; Darling-Hammond, L.; Haertel, E. e Thomas, E. (2010) "Value-added modelling of teacher effectiveness: an exploration of stability across models and contexts", *Education Policy Analysis Archives*, Vol. 18, n. 23, pp. 1-22

Nilsson, O. and Loughran, J. (2011) "Exploring the Development of Pre-Service Science Elementary Teachers' Pedagogical Content Knowledge". *J Sci Teacher Educ*, No. 23, pp.699-721

Noel, A. (2014) "Teacher Performance Assessment (edTPA): An Instructor's Development and Evaluation of an Embedded Signature Assessment in an Early Childhood Literacy Course", *Journal of Early Childhood Teacher Education*, Vol. 35, n. 4, pp. 357-372

Norton, A. McCloskey, A. and Hudson, R.A. (2011) "Prediction assessments: Using video-based predictions to assess prospective teachers' knowledge of students' mathematical thinking". *J Math Teacher Educ*, No. 14, pp. 305-325

Nusche, D., et al. (2014) , *OECD Reviews of Evaluation and Assessment in Education: Netherlands*, OECD Publishing.

Oliver, D.F. and Huffman, J.B. (2016) "Professional learning community process in the United States: conceptualization of the process and district support for schools", *Asia Pacific Journal of Education*, Vol. 36, No. 2, pp.301-317



Ozel, Z.E.Y. and Ozel, S. (2013) "Mathematics teacher quality: in distribution and relationship with student achievement in Turkey", *Asia Pacific Educ. Rev.*, No. 14 pp. 231-242

Pakarinen, E.; Lerkkanen, M.; Poikkeus, A.; Kiuru, N.; Siekkinen, M.; Rasku-Puttonen, H. e Nurmi, J. (2010) "A Validation of the Classroom Assessment Scoring System in Finnish Kindergartens", *Early Education and Development*, Vol. 21, n. 1, pp. 95-124

Popp, P.; Grant, L. e Stronge, J. (2011) "Effective Teachers for At-Risk or Highly Mobile Students: What are the Dispositions and Behaviors of Award-Winning Teachers?", *Journal of Education for Students Placed at Risk (JESPAR)*, 16:4, 275-291

Ramezani-Monfared, N. Shahvarani, A. and Behzadi, M.H. (2015) "The Impact of Mathematics Teachers; Effectiveness on Students' Learning in the Two Realms of: Knowledge and Understanding". *Mathematics Education Trends and Research*, No. 1 pp.35-42

Ray, A.; McCormack, T. e Evans, H. (2009) "Value Added in English Schools", *Education Finance and Policy*, Vol. 4, n. 4, pp. 415-438

Selkrig, M. Keamy, K. (2017) "Creative pedagogy: a case for teachers' creative learning being at the centre", *Teaching Education*, Vol. 28, n. 3, pp. 317-332

Sonmark, K. et al (2017) "Understanding teacher's pedagogical knowledge", *OECD Education Working Papers*, n. 159, OECD Publishing, Paris.

Strong, M. Gargani, J. Hacifazlioglu, O. (2011) "Do We Know a Successful Teacher When We See One? Experiments in the Identification of Effective Teachers", *Journal of Teacher Education*, 62;4, pp.367-382

Thi Mai, H. e Hall, C. (2017) "Changing views of teachers and teaching in Vietnam", *Teaching Education*, Vol. 28, n. 3, pp. 244-256

Tucker, B.M. (2013) "Student evaluation to improve the student learning experience: an Australian university case study", *Educational Research and Evaluation*, Vol. 19, No. 7, pp.615-627

Tschannen-Moran , M., Hoy, A.W. and Hoy, W.K. (1998) "Teacher Efficacy: Its Meaning and Measure", *Review of Educational Research*, Vol. 68, No. 2, pp. 202-248

Walkington C., Arora, P., Ihorn, S., Gordon, J., Walker, M., Abraham, L. e Marder, M. (2011) *Development of the UTeach Observation Protocol: A Classroom Observation Instrument to Evaluate Mathematics and Science Teachers from the UTeach Preparation Program* | Request PDF. Available from:  
[https://www.researchgate.net/publication/265478444\\_Development\\_of\\_the\\_UTeach\\_Obse rvation\\_Protocol\\_A\\_Classroom\\_Observation\\_Instrument\\_to\\_Evaluate\\_Mathematics\\_and\\_Science\\_Teachers\\_from\\_the\\_UTeach\\_Preparation\\_Program](https://www.researchgate.net/publication/265478444_Development_of_the_UTeach_Obse rvation_Protocol_A_Classroom_Observation_Instrument_to_Evaluate_Mathematics_and_Science_Teachers_from_the_UTeach_Preparation_Program)

Weems, D.M. and Rogers C.B.H. (2010) "Are US teachers making the grade? A proposed framework for teacher evaluation and professional growth", *Management in Education*, Vol. 24 Iss. 1 pp. 19-24

Welsh, M.E. (2011) "Measuring Teacher Effectiveness in Gifted Education: Some Challenges and Suggestions", *Journal of Advances Academics*, Vol. 22 No. 5 pp. 750-770

Westrick, J.M and Morris, G.A. (2016) "Teacher education pedagogy: disrupting the apprenticeship of observation", *Teaching Education*, Vol. 27 No. 2 pp. 156-172

Wijaya, A. Heuvel-Panhuizen, M. and Doorman, M. (2015) "Teacher's teaching practices and beliefs regarding context-based tasks and their relation with students difficulties in solving these tasks" *Math Ed Res J*, No. 27 pp. 637-662

Wilkerson, J.R. (2015) "The Need for Instrumental Sensitivity and Construct Clarity in PACT: A Commentary on "Examining the Internal Structure Evidence for the Performance Assessment for California Teachers", *Journal of Teacher Education*, Vol. 66 No. 2 pp. 184-192

Winterbottom, C. and Winterbottom, S.S. (2017) "Social Justice: A Case-Study Examining the Influence of Primary Headteachers in Two Manchester Schools", *Early Childhood Educ. J*, No. 45 pp. 755-768

Whitehurst, G.J.R. Chingos, M.M. and Lindquist, K.M. (2015) "Getting? Classroom? Observations? Right", *Education Next*, Vol. 15 Iss. 1

Whittaker, J.E.V. Willford, A.P. Carter, L.M Vitiello, V.E. Hatfield, B.E. (2018) "Using a Standardized Task to Assess the Quality of Teacher-Child Dyadic Interactions in Preschool", *Early Education Development*, Vol. 29 No. 2 pp. 266-287

Wright, P.M. and Craig M.W. (2011) "Tool for Assessing Responsibility-Based Education (TARE): Instrument Development, Content Validity, and Inter-Rater Reliability", *Measurement in Physical Education and Exercise Science.*, Rev. 15 No. 3 pp. 204-219

Yildirim, A. and Kasapoglu, K. (2015) "Teacher's perceptions of constructivist curriculum change as a predictor of their perceptions of the implementation of constructivist teaching", *Asia Pacific Edu.* No. 16 pp. 565-577

Yin, H. Han, J. and Lu, G. (2017) "Chinese tertiary teacher's goal orientations for teaching approaches: the mediation of teacher engagement", *Teaching in Higher Educatio.*, Vol. 22 No. 7 pp. 766-784

Yoon, T. (2013) "Strengthening Teacher Education Program: Keys to Develop through Teacher Assessment", *Journal of Education and Learning*, Vol. 8 No. 2 pp. 91-97

Yuan, R. and Zhang, J. (2016) "Promoting Teacher Collaboration Through Joint Lesson Planning: Challenges and Coping Strategies" *Asia Pacific Edu Res*, No. 25 pp. 817-826

Yuan, K. (2015) "A Value-Added Study of Teacher Spillover Effects Across Four Core Subjects in Middle Schools", *Education Policy Analysis Archives*, Vol. 23 No. 38

Zach, S. Harari, I. and Harari, N (2012) "Changes in teaching efficacy of pre-service teachers in physical education", *Physical Education and Sport Pedagogy*, Vol. 17 No. 15 pp. 447-462

Zerihun, Z. Beishuizen, J. and Van Os, W. (2012) "Student learning experience as indicator of teaching quality", *Educ Asse Eval Acc*, No. 24 pp. 99-111

Zheng, C. Fu, L. and He, P. (2012) "Development of an Instrument for Assessing the Effectiveness of Chemistry Classroom Teaching", *Journal of Science Education*, Vol. 23 No. 2 pp. 267-279

## ANEXO – DETALHAMENTO DOS CRITÉRIOS DE AVALIAÇÃO E QUESTIONÁRIOS

### 1. CRITÉRIOS DE APOIO A AVALIAÇÕES DE QUALIDADE DE DARLING-HAMMOND ET AL (2013)

1. Avaliação de Habilidades Cognitivas de Alta-Ordem que permitem os estudantes transferirem seus aprendizados para novas situações e problemas
2. Avaliação de Alta-Fidelidade de Habilidades Críticas tais como usadas no mundo real ao invés de proxies artificiais (medidas de avaliação de desempenho de habilidades orais, escritas e de comunicação, colaboração, pesquisa, experimentação assim como sobre o uso de novas tecnologias)
3. Avaliações que são lastreadas internacionalmente a partir de avaliações pré-existentes em países de referência educacional, com foco nas atividades que representam e níveis de desempenho esperados
4. Uso de itens que têm valor educacional incluindo testes que são definidos para captar conceitos ensinados específicos ao invés de depender de testes que podem ser influenciados por experiências fora da sala de aula. Em particular, avaliar se os testes oferecem *insights* sobre como os estudantes aprendem
5. Avaliações que sejam válidas, confiáveis e justas que devem captar as habilidades dos estudantes que se tem a intenção de medir, que sejam livres de vieses e desenhadas para reduzir obstáculos desnecessários para sua validade com qualidades positivas para a qualidade da instrução.

### 2. ABORDAGEM ‘APRENDENDO A NOTAR’ (LEARNING TO NOTICE FRAMEWORK)

	Nível 1	Nível 2	Nível 3	Nível 4
	Linha de Base	Mista	Focada	Ampla
O que os professores notam	Contempla o ambiente de toda a classe, a pedagogia	Trata prioritariamente da pedagogia de ensino, mas inclui	Contempla unicamente o pensamento dos alunos	Relaciona o pensamento dos alunos e as estratégias de

	do professor e o aprendizado	também o pensamento e comportamento dos alunos		ensino dos professores
--	---------------------------------	---	--	---------------------------

Fonte: van Es, 2011, p 139

### 3. SURVEY DE PRÁTICAS REFLEXIVAS E COLABORATIVAS DE CAMBURN

Camburn, E. (2010) "Learning Opportunities as a site for reflective practice: an exploratory study". *American Journal of Education*, vol. 116, n. 4, pp. 463-489

#### Escala de Práticas Reflexivas

As minhas oportunidades de aprendizado este ano....	Discordo fortemente (%)	Discordo (%)	Concordo (%)	Concordo fortemente (%)
Deram um retorno útil as minhas práticas pedagógicas				
Permitiram-me focar em um problema sobre um período extenso de tempo				
Possibilitaram que buscasse informações adicionais sobre outros professores, algum instrutor líder, etc				
Levaram-me a pensar sobre algum aspecto das minhas práticas pedagógicas de algum ângulo novo				
Fizeram-me prestar mais atenção em ações pedagógicas particulares que eu realizo em sala-de-aula				

Me deram oportunidades de trabalhar em aspectos de minhas aulas que eu estou tentando desenvolver				
Me deram conhecimento/informação que é muito útil para mim em sala-de-aula				
Levaram-me a tentar novas práticas educacionais em sala-de-aula				

### Escala Colaborativa

<b>Durante esse ano escolar, quantas vezes você fez o seguinte:</b>	<b>Nunca (%)</b>	<b>1-2 vezes por ano (%)</b>	<b>3-5 vezes por ano (%)</b>	<b>6-10 vezes por ano (%)</b>	<b>Mais de 10 vezes por ano (%)</b>
Viu outro professor dar aula e deu algum feedback para ele ou ela					
Algum outro professor o observou e lhe deu um feedback					
Assistiu qualquer instrução dada por outro professor					
Examinou/mudou o escopo/sequência de algum tópico curricular junto com outros colegas					



Aprendeu a adotar e usar estratégias de instruções de grupo com colegas					
Desenvolveu unidades temáticas para instruções de integração no currículo junto com outros colegas					
Examinou o alinhamento de materiais curriculares e avaliações com colegas					
Esclareceu padrões de aprendizado através de discussões e análises com colegas					

### Escala de Trabalho com Experts educacionais

Durante esse ano escolar, quantas vezes aconteceu o seguinte:	Nunca (%)	1-2 vezes por ano (%)	3-5 vezes por ano (%)	6-10 vezes por ano (%)	Mais de 10 vezes por ano (%)
Um avaliador (ou como eles chamam 'instructional leader') estudou o trabalho de meus					

estudantes e deu sugestões de como melhorá-los					
Um avaliador me deu feedback sobre o meu uso de materiais curriculares					
Eu tive instruções de um avaliador (ou de um líder de instruções, o que nesse contexto não é a mesma coisa)					
Um avaliador me deu feedback sobre como melhorar minhas técnicas de ensino					

#### 4. INSTRUÇÕES PARA SIMULAÇÕES DE VÍDEO

Fonte: Amador, J. (2017) "Preservice teachers' vídeo simulations and subseqüente noticing: a practice-based method to prepare mathematics teachers" Research in Mathematics Education, vol. 19, n. 3, pp. 217-235.

Essas instruções foram planejadas para o desenvolvimento de conteúdos de matemática mas podem ser aplicadas para qualquer outra disciplina.

"A sua tarefa é desenhar, criar e filmar um cenário de uma classe hipotética envolvendo uma discussão matemática baseada no seguinte critério:

2.MD.C.8 do "Common Core State Standards for Mathematics" (poderíamos traduzir isso como os padrões curriculares básicos para matemática do Estado): resolução de problemas descritivos envolvendo notas de dinheiro, 50 centavos, 25 centavos, dez centavos e

centavos usando os símbolos correspondentes. Por exemplo, se você tem 2 moedas de 10 centavos e 3 centavos, quantos centavos você tem? (CCSSI, 2010, p. 20)

1. Escreva um script (palavra por palavra) de uma discussão completa em sala-de-aula que exemplificasse as melhores circunstâncias possíveis para o aprendizado matemático.
  - a. O seu script deve ser baseado em um problema matemático real, por sua vez fundamentado nos padrões curriculares básicos (você deve criar esse problema e selecionar o nível de avaliação)
  - b. O seu script deve incorporar todos os cinco movimentos de fala segundo Chapin et al (2009) [Chapin, S., O'Connor, C. & Anderson, N. (2009) *Classroom discussions: using math talk to help students learn*. Sausalito, CS: Math Solutions.
  - c. Torne o script tão realista quanto possível, antecipe como os estudantes de verdade responderiam
  - d. Tome decisões sobre o que deve ser incluído no script com base no livro de Chapin et al (2009)
  - e. O seu script deve ter não menos que 2 minutos e não mais que 6 minutos de filmagem
  - f. Seja criativo e lembre que aprender matemática [assim como qualquer outra disciplina] é o foco de uma discussão matemática
  - g. Uma cópia final do seu registro de grupo será coletada
  - h. Baseado em suas leituras, incorpore evidências de entendimento de um de seus elementos do script: estratégias de avaliação, ensino para todos, integração de tecnologias
2. Encontre um lugar na sua escola de ensino fundamental onde você possa filmar sua discussão sem perturbar nenhuma sala-de-aula
  - a. Filme o seu script, designando um dos membros do seu grupo como o diretor de vídeo. Todos os outros participantes do seu grupo devem aparecer no filme e devem fazer parte da conversação
  - b. Pode ser que você precise gravar alguma cena várias vezes para poder produzir um vídeo de qualidade

- c. Esteja preparado para justificar os conteúdos do seu vídeo e porque é um exemplo representativo da uma discussão de sala-de-aula
3. Prepare para compartilhar a sua simulação com seus colegas e critique as simulações dos seus colegas. Esta é a sua oportunidade de demonstrar o seu entendimento e aplicabilidade dos 5 movimentos de fala de Chapin et al, 2009. Pense cuidadosamente como poderiam ser incorporados em uma lição.

## **5. RUBRICAS DE AVALIAÇÃO DE PREVISÕES DE VÍDEO DE NORTON, MCCLOSKEY & HUDSON (2011)**

Norton, A., McCloskey, A. & Hudson, R. (2011) "Prediction assessments: using vídeo-based predictions to assess prospective teacher's knowledge of students' mathematical thinking" *Journal of Math Teacher Education*, vol. 14, pp. 305-325.

Conhecimento do conteúdo:

0: se resolveu o problema incorretamente

1: se resolveu o problema corretamente

Modelo:

0: não usa evidência (descrições das ações dos estudantes ou afirmações) para descrever o que ou como o estudante está pensando

1(a): usa evidência para fundamentar uma explicação do que o estudante conhece ou pensa, mas não *como* o estudante articula o seu pensamento

1(b): explica como o estudante está pensando, mas não dá nenhuma evidência explícita para fundamentar sua explicação

2: usa evidência para fundamentar uma explicação razoável de *como* o estudante articula o seu pensamento

Previsão (precisão e detalhe):

0: não faz nenhuma previsão relevante para a situação

1: faz uma previsão imprecisa com algum detalhe relevante para a situação, mas não o suficiente para ver de modo não ambíguo o que o estudante teria feito ou dito em resposta a uma determinada tarefa ou questão

2(a): faz uma previsão imprecisa, mas com detalhe relevante suficiente para ver o que o estudante teria feito ou dito em resposta a uma determinada tarefa ou questão

2(b): faz uma previsão que poderia ser correta, mas permanece demasiado vaga

3(a): faz uma previsão precisa com algum detalhe relevante para a situação, mas não suficientemente sem ambiguidades para prever o que o estudante faria ou diria em resposta a uma determinada tarefa ou questão

3(b): faz múltiplas previsões, uma das quais descreve com precisão o que o estudante diria ou faria

4: faz uma previsão precisa com detalhe suficiente para prever o que o estudante diria ou faria em resposta a uma determinada tarefa ou questão

Uso do Modelo:

0: não há evidência (ou há evidência contrária) que o professor usou uma explicação do pensamento dos estudantes (modelo) para elaborar qualquer uma das previsões

1: existe alguma evidência que o professor usou um modelo para elaborar alguma das previsões

2: o professor claramente usou um modelo para formar a maior parte ou todas as previsões

## 6. CRITÉRIOS 'INFORM' PARA ESCOLHA DE OBSERVAÇÕES DE CAMPO

Hoyt, K. & Terantino, J. (2015) "Rethinking Field Observations: strengthening teacher education through INFORM" Action in Teacher Education, vol. 37, n. 3, pp. 209-222.

O quê	Quem	Onde
1a. Selecionar escolas e professores	Instrutor de métodos	parceiros
1b. Agendar visitas as escolas e compartilhar o guia de observações de classe	Instrutor de métodos e professores de classe	Por email
2a. Realização das práticas usando o guia de observações	Estudantes	Tema de casa: ver vinhetas de vídeo-aulas
2b. 'Debriefing' usando o guia	Estudantes e Instrutor de métodos	Classes da universidade (ou outras classes)
3. Condução das observações de campo	Estudantes e Instrutor de métodos	Escolas
4. Condução dos 'debriefs' das observações	Estudantes, instrutor de métodos e professores de classe	Escolas
5. Escrita das análises das observações de campo (ciclo de feedback com reflexões amplas)	Estudantes e instrutor de métodos	Tema de casa

### FORMULÁRIO INFORM – GUIA DE OBSERVAÇÃO

Candidato \_\_\_\_\_ a \_\_\_\_\_ professor:

\_\_\_\_\_ Data: \_\_\_\_\_

Escola \_\_\_\_\_ (lugar \_\_\_\_\_ da \_\_\_\_\_ observação):

Professor \_\_\_\_\_ da \_\_\_\_\_ classe:

\_\_\_\_\_

Classe: \_\_\_\_\_ Idade \_\_\_\_\_ dos  
estudantes: \_\_\_\_\_

## Instruções

Para a sua referência para completar o seu Relatório Inform, é recomendável que você tome notas das lições que você vê à medida que elas vão sendo apresentados. Preste atenção a todas atividades observadas e note se elas são **centradas no professor** ou **centradas nos alunos**. O que você observou que poderia ser caracterizado como uma **instrução baseada em padrões**? Também tome nota da quantidade de tempo gasto em cada atividade.

Anote os itens dispostos nas seções I-III, como sugerido abaixo. Esses itens refletem o conteúdo da *Reflexão Escrita INFORM* que você completará e submeterá seguindo cada visita de classe.

### I. Tomada de Decisão dos Professores

Tomada de decisão dos professores (no seu planejamento): O que você julga que são os **resultados de aprendizagem esperados para os estudantes** para a lição observada? Escreva dois ou três objetivos antecipados ou resultados de aprendizagem.

Tomada de decisão dos professores (no momento da aula): Você pode identificar um incidente no qual o professor tomou uma decisão para fazer alguma coisa (que possivelmente não estivesse incluída no plano de aula original) baseada em uma situação não-antecipada?

### II. Uso de Linguagem Direcionada (do inglês, "Target Language Use")

Uso do professor da linguagem direcionada: Que porcentagem da linguagem direcionada você pensa que foi usada pelo professor? A abordagem do professor lhe pareceu ser efetiva?

Uso dos estudantes da linguagem direcionada: Quanto e de que modos os estudantes usaram a linguagem direcionada? Foi feita de modo significativo? Discuta os **modos de comunicação** que foram utilizados pelos estudantes (interpessoal, interpretativo, e expositivos). Anote exemplos específicos.

### III. Uso da Tecnologia

Que tipo de tecnologia foi utilizada nas aulas, se utilizada de todo? Por quê você pensa que o professor usou esse tipo de tecnologia? Você acha que ela foi usada de modo significativo durante a lição? Se nenhuma tecnologia foi utilizada, você poderia sugerir alguma que pudesse ter sido utilizada?

## FORMULÁRIO INFORM – RELATÓRIO ESCRITO REFLEXIVO

Candidato \_\_\_\_\_ a \_\_\_\_\_ professor:  
\_\_\_\_\_ Data: \_\_\_\_\_

Escola \_\_\_\_\_ (lugar \_\_\_\_\_ da \_\_\_\_\_ observação):  
\_\_\_\_\_

Professor \_\_\_\_\_ da \_\_\_\_\_ classe:  
\_\_\_\_\_

Classe: \_\_\_\_\_ Idade \_\_\_\_\_ dos  
estudantes: \_\_\_\_\_

Instruções



Revise suas anotações tomadas durante a visita a classe INFORM e escreva uma resposta narrativa de um ou dois parágrafos para cada um dos itens abaixo. (Use este documento como um roteiro para o seu relatório escrito e componha suas respostas diretamente seguindo cada item.)

## I. Tomada de Decisão dos Professores

Tomada de decisão dos professores (no seu planejamento): O que você julga que são os **resultados de aprendizagem esperados para os estudantes** para a lição observada? Escreva dois ou três objetivos antecipados ou resultados de aprendizagem e o que você notou durante a classe que seria evidência da realização desses objetivos.

Tomada de decisão dos professores (no momento da aula): Você pode identificar um incidente no qual o professor tomou uma decisão para fazer alguma coisa (que possivelmente não estivesse incluída no plano de aula original) baseada em uma situação não-antecipada? Como a tomada de decisão do professor foi discutida durante o *debriefing* e o que você entendeu dessa discussão?

## II. Uso de Linguagem Direcionada (do inglês, "Target Language Use")

Uso do professor da linguagem direcionada: Que porcentagem da linguagem direcionada você observou sendo usada pelo professor? A abordagem do professor lhe pareceu ser efetiva? A linguagem direcionada foi usada durante o *debriefing*? Se sim, como foi usada?

Uso dos estudantes da linguagem direcionada: Quanto e de que modos os estudantes usaram a linguagem direcionada? Foi feita de modo significativo? Discuta os **modos de comunicação** que foram utilizados pelos estudantes (interpessoal, interpretativo e expositivos). Anote exemplos específicos.

### III. Uso da Tecnologia

Que tipo de tecnologia foi utilizada nas aulas, se utilizada de todo? Por quê você pensa que o professor usou esse tipo de tecnologia? Você acha que ela foi usada de modo significativo durante a lição? Se nenhuma tecnologia foi utilizada, você poderia sugerir alguma que pudesse ter sido utilizada?

### IV. Reflexão Pessoal

O que você aprendeu das observações e do *debriefing*? Você observou alguma coisa que você definitivamente nunca fará em suas próprias aulas? Houve alguma estratégia de observação ou atividade que você gostaria de incorporar em suas próprias classes? O que você ganhou pessoalmente do *debriefing*? Seja específico e explique.

## **7. QUESTIONÁRIO DE PERCEPÇÃO DOS ESTUDANTES – TRIPOD (PARA ESTUDANTES DO INÍCIO DO ENSINO FUNDAMENTAL – V16)**

<https://www.staffordschools.org/cms/lib/NJ01001734/Centricity/Domain/24/Student%20Survey.pdf>

Caro estudante

Esse questionário é sobre as suas aulas na escola. Depois de cada linha, por favor marque uma das situações, 'sim', 'talvez' ou 'não' para mostrar o quanto você concorda ou não com cada opção.

Por favor, marque o que você realmente pensa, mesmo se é diferente do que outras pessoas pensam. Se houver uma linha que você não entender ou que não quiser responder, por favor, deixe em branco.

Obrigado pelo seu tempo!

1. Essa classe é um lugar feliz para eu estar.

não     talvez     sim

2. Eu me esforço muito para entender tudo o que o meu professor ensina.

não     talvez     sim

3. Nessa classe nós aprendemos muito quase todos os dias.

não     talvez     sim

4. Os meus colegas de classe 'pegam no meu pé' e ferem os meus sentimentos.

não     talvez     sim

5. Os meus colegas de classe agem do modo que o meu professor deseja.

não     talvez     sim

6. Eu tento tanto quanto posso aprender na escola.

não     talvez     sim

7. O meu professor escuta com atenção quando as crianças falam.

não     talvez     sim

8. Nessa classe, o aprendizado é devagar e não é muito divertido. Você concorda?

não     talvez     sim

9. Nessa classe, eu desisto e paro de tentar quando o trabalho é difícil.

não     talvez     sim

10. Muitos dias eu fico muito sonolento na escola.

não     talvez     sim

11. Alguns colegas de classe debocham daquelas crianças que erram.

não     talvez     sim

12. Eu leio em casa quase todos os dias.

( ) não ( ) talvez ( ) sim

13. Eu gosto das coisas que estamos aprendendo nessa classe.

( ) não ( ) talvez ( ) sim

14. Eu gosto de como o meu professor me trata quando eu preciso de ajuda.

( ) não ( ) talvez ( ) sim

15. O meu professor é muito bom em explicar coisas.

( ) não ( ) talvez ( ) sim

16. Mesmo quando alguma coisa é difícil para mim, o meu professor consegue que eu entenda. ( ) não ( ) talvez ( ) sim

17. A nossa classe está sempre ocupada e não desperdiça tempo.

( ) não ( ) talvez ( ) sim

18. O meu professor toma tempo para nos ajudar a lembrar o que aprendemos.

( ) não ( ) talvez ( ) sim

19. O meu professor garante que eu tente fazer o meu melhor.

( ) não ( ) talvez ( ) sim

20. O meu professor é gentil comigo quando eu faço perguntas.

( ) não ( ) talvez ( ) sim

21. Em nossa classe, está ok de parar de tentar.

( ) não ( ) talvez ( ) sim

22. Quando o professor está nos ensinando, nos pergunta se entendemos o que ensina.

( ) não ( ) talvez ( ) sim

23. Para nos ajudar a lembrar, o meu professor fala sobre coisas que já aprendemos.

não  talvez  sim

24. O meu professor quer que eu explique minhas respostas – por que eu penso o que eu penso.  não  talvez  sim

25. Eu gosto de estar na escola.

não  talvez  sim

26. Aprender é muitas vezes difícil para mim.

não  talvez  sim

27. Você é menino ou menina?

menino  menina

28. Há um computador na sua casa?

não  sim, há um  sim, há mais de um

29. A sua família fala inglês em casa?

sempre  algumas vezes  nunca

30. Com que adultos você vive? (Marquem todos que forem o seu caso)

minha mãe  meu pai  minha avó  meu avô

minha madrasta  meu padrasto  outros adultos

31. Há uma televisão no quarto onde você dorme?

não  sim

32. Em que séries você está?

pré-escola  primeiro ano  segundo ano  outro

## 8. CRITÉRIOS DE AVALIAÇÃO DE CRIATIVIDADE DE ESTUDANTES EM AULAS DE MATEMÁTICA DE COLES AND SCOTT

Coles, A. & Scott, H. (2015) "Planning for the Unexpected in the Mathematics Classroom: an account of teacher and student change". *Research in Mathematics Education*, vol. 17, n. 2, pp. 128-147

1. Identificar se os estudantes fazem suas próprias perguntas
2. Avaliar se os estudantes seguem as suas próprias linhas de duvida
3. Verificar se os estudantes escolhem os seus próprios métodos de representação
4. Identificar se os estudantes nota padrões naqueles exercícios matemáticos que estão fazendo
5. Avaliar se os estudantes conseguem fazer previsões ou pelo menos algumas conjecturas sobre os problemas discutidos

## 9. PROTOCOLO DE OBSERVAÇÃO DE SALA-DE-AULA DE STRONG ET AL (2011)

Strong, M., Gargani, J. & Hacifazlioglu, O. (2011) "Do we know a successful teacher when we see one? Experiments in the identification of effective teachers" *Journal of Teacher Education*, vol. 62, n. 4, pp. 367-382.

### INSTRUÇÕES PARA AVALIADORES

"Você verá sete clips de 2 minutos cada de professores do ensino fundamental dando uma lição de matemática. Não há interrupção entre os clips. Alguns desses professores têm classes que tem um desempenho acima da média ano após ano. Outros têm classes com desempenho moderado/médio no aprendizado dos alunos. Para cada clip, por favor, circule 'sim' na planilha de notas se você pensa que o professor pertence ao grupo acima da média ou 'não' se você pensa que o professor pertence ao grupo que está na média. Assegure-se que você associa a sua nota ao número correto do professor. Você pode mudar de opinião a qualquer momento, se você desejar.

### PROTOCOLO DE *DEBRIEF*

As Notas

1. Alguns dos professores você classificou como de alto desempenho. Que lhe chamou atenção sobre esses professores que fizeram você classificá-los nesse grupo? (forneça *screen shots* dos professores). Você poderia dizer algo sobre eles como um grupo e depois falar sobre eles individualmente? (tente detalhar a informação ou dar exemplos ilustrativos se forem usadas descrições como 'carisma', 'presença', 'engajamento', etc)
2. Aqui estão os outros professores. Poderia falar sobre o que você viu nos clips deles que afetaram o seu julgamento sobre eles? (detalhe quando necessário)
3. In geral, pensando sobre todos esses clips, que critérios você acabou usando para ordenar/classificar os professores? (use os sistemas de significado subjacentes, valores, propósitos da educação e exemplos ilustrativos)
4. Por quê você confia nesses critérios?
5. Se você pudesse ter tido outro pedaço de informação para tomar a sua decisão, o que teria escolhido? Por quê?
6. O que você acha que é o valor de usar os critérios que você mencionou?
7. Existem quaisquer limitações no uso desses critérios? Se sim, quais seriam?

#### A Tarefa

1. De que modos pode-se dizer que o exercício foi fácil ou difícil? Por favor explique e dê exemplos.
2. Se você tivesse olhado somente um minuto para cada professor, será que você teria chegado aos mesmos resultados? E se tivessem sido 30 segundos? Você acha que teria mudado o seu julgamento se você tivesse olhado os professores por 10 minutos ou mesmo por uma lição completa? (Explique porque sim ou não)
3. E se eu lhe dissesse que seus julgamentos foram 100% corretos –isso lhe surpreenderia? A que você atribuiria o seu êxito? (Forneça explicações profundas, fontes ou experiências relevantes)
4. E se eu lhe dissesse que seus julgamentos foram na sua maior parte errados –isso lhe surpreenderia? A que você atribuiria sua falta de acertos?
5. O que você supõe que isso lhe diz sobre dar aulas?
6. O que você supõe que isso lhe diz sobre a qualidade dos professores?
7. Existe alguma coisa a mais que esse exercício lhe faz pensar que você gostaria de mencionar?

## 10. TEMPLATE CoRe INCLUINDO A ESCALA DE AUTO-AVALIAÇÃO

Nisson, P. & Loughran, J. (2012) "Exploring the Development of Pre-Service Science Elementary Teachers' Pedagogical Content Knowledge" *Journal of Science Teacher Education*, vol. 23, pp. 699-721

Conteúdo:	Grande ideia A	Grande ideia B	Grande ideia C	O quanto significativo ou importante é essa questão para mim quando eu penso sobre minhas aulas? (marque em uma escala de 1 (baixo) a 10 (alto))	O quanto confiante me sinto quando eu respondo essa questão? (marque em uma escala de 1 (baixo) a 10 (alto))
Idade da criança:					
O que você pretende que os seus estudantes aprendam sobre essa ideia?					
Por quê é importante que os seus estudantes saibam isso?					
O que mais você sabe sobre essa ideia (que você não quer que os seus estudantes saibam ainda)?					
Que dificuldades/limitações estão presentes ao ensinar essa ideia?					
Qual é o seu conhecimento sobre o processo de pensamento dos seus estudantes que					



influencia o jeito de ensinar essa ideia?					
Que outros fatores lhe influenciam no modo de você ensinar essa ideia?					
Que procedimentos de ensino você usa e quais são as razões particulares para usá-los ao trabalhar essa ideia?					
Que modos específicos você tem para descobrir se os estudantes entenderam ou estão confusos com suas explicações?					

## 11. CINCO DIMENSÕES do PLC – “PROFESSIONAL LEARNING COMMUNITY’ E PROTOCOLO DE ENTREVISTA

Olivier, D. & Huffman, J. (2016) “Professional Learning Community process in the United States: conceptualization of the process and district support for schools”. *Asia Pacific Journal of Education*, vol. 36, n. 2, pp. 301-317

1. Liderança compartilhada e apoiadoras: os administradores escolares compartilham poder, autoridade e tomada de decisão, promovendo e nutrindo lideranças
2. Valores compartilhados e visão: os funcionários compartilham visões que tem um foco no aprendizado dos alunos e apoiam normas de comportamento que guiam decisões sobre ensino e aprendizado
3. Aprendizagem coletiva e aplicações: os funcionários compartilham informação e trabalham coletivamente para planejar, resolver problemas e melhorar as oportunidades de aprendizado
4. Prática pessoal compartilhada: os pares se encontram e observam um ao outro para dar feedback sobre práticas educacionais, para ajudar no aprendizado dos alunos e para aumentar a capacidade humana

5. Condições de suporte: relações incluem respeito, confiança, normas de pesquisa crítica e melhorias e relações positivas, de cuidado entre toda a comunidade escolar. Estruturas incluem sistemas (e.g. comunicação e tecnologia) e recursos (e.g. pessoal, prédios, tempo, fiscal e material) que possibilitam aos funcionários enfrentar e examinar as práticas e resultados dos estudantes.

## PROTOCOLO DE ENTREVISTA

### Avaliação do Apoio a Comunidades de Aprendizado Profissional

#### Liderança compartilhada e apoiadora

Descreva os modos pelos quais a sua cidade (poderia ser também sua 'secretaria de educação) facilita uma liderança compartilhada e apoiadora dentro de suas escolas.

#### Valores compartilhados e visão

Descreva como a sua cidade apoia o desenvolvimento de valores compartilhados e visão em suas escolas.

#### Aprendizagem coletiva e aplicações

Descreva como a sua cidade apoia aprendizado coletivo e aplicações em suas escolas.

#### Prática pessoal compartilhada

Descreva como a sua cidade apoia os professores no compartilhamento de suas práticas de sala-de-aula e estratégias em suas escolas.

#### Condições de suporte – estruturas e relações

Descreva como a sua cidade fornece condições de apoio em relação as estruturas de suas escolas.

Descreva como a sua cidade fornece condições de apoio em termos de relações com as escolas.

Foco dos dados

Descreva como a sua cidade apoia o uso de dados para melhorar o ensino e aprendizado nas suas escolas.

Desafios

Descreva quaisquer desafios que poderiam existir dentro da sua cidade relacionados ao apoio do processo PLC em suas escolas.

Questão Final

Estamos chegando ao fim de nossa entrevista. Há algo que você gostaria de adicionar sobre o que sua cidade poderia iniciar, parar ou continuar no processo PLC dentro de sua escola?

## **12. TKS– “TEACHER KNOWLEDGE SURVEY” OU PESQUISA DO CONHECIMENTO DOS PROFESSORES DA OCDE (2017)**

Sonmark, K., Revai, N., Gottschalk, F., Deligiannidi, K. e Burns, T. (2017) “Understanding Teachers’ pedagogical knowledge – Report on an international pilot study”. OECD Education Working Papers n. 159, OECD.

(a traduzir se julgado conveniente, 9 páginas da p. 126-134)

## **Oportunidades de aprendizado**

Tabela II.5.1 Oportunidades de aprender as principais dimensões do conhecimento: processos de instrução; processos de aprendizado; avaliação

Tabela II.5.2 Qualidade de oportunidades para aprender: qualidade da instrução; demandas; agência dos estudantes

Tabela II.5.3 Atividades de pesquisa: engajamento na pesquisa, e com a pesquisa

Tabela II.5.4 Colaboração profissional

Tabela II.5.5 Experiência de campo: quantidade de experiências; escopo das experiências; apoio profissional

Tabela II.5.6 Avaliação do desempenho dos professores educadores

Tabela II.5.7 Práticas de ensino dos professores educadores

## **Qualidade da prática de ensino**

Tabela II.6.1 Qualidade da prática de ensino

Tabela II.6.2 Escala de auto-eficácia do professor

Tabela II.6.3 Escala das motivações para ensinar

Tabela II.6.4 Orientações de objetivos

Tabela II.6.5 Responsabilidade própria do professor

Tabela II.6.6 Orientações intrínsecas

Tabela II.6.7 Comprometimento com o ensino

### **13. RESUMO DOS ASPECTOS AVALIADOS NO INSTRUMENTO 'EVALUACIÓN DOCENTE' NO MÉXICO**

Perfil Docente	Etapas		
	Informe de responsabilidades profesionales	Proyecto de Enseñanza	Examen de conocimientos didácticos y curriculares
<b>Dimensión 1. Un docente que conoce a sus alumnos, sabe cómo aprenden y lo que deben aprender</b>	No se aborda en el instrumento	No se aborda en el instrumento	1.1.1, 1.1.2, 1.2.1, 1.2.2, 1.2.3, 1.2.4, 1.3.1, 1.3.2.
1.1 Conoce los procesos de desarrollo y de aprendizaje infantiles.			
1.2 Domina los propósitos educativos y los contenidos escolares de la Educación Primaria.			
1.3 Explica los referentes pedagógicos y los enfoques didácticos del currículo vigente.			
<b>Dimensión 2. Un docente que organiza y evalúa el trabajo educativo, y realiza una intervención didáctica pertinente</b>	No se aborda en el instrumento	2.1.1, 2.1.2, 2.1.3, 2.1.4, 2.2.1, 2.2.2, 2.2.3, 2.2.4, 2.2.5, 2.2.6, 2.3.1, 2.3.2, 2.4.1, 2.4.2, 2.4.3.	No se aborda en el instrumento
2.1 Organiza su intervención docente para el aprendizaje de sus alumnos.			
2.2 Desarrolla estrategias didácticas para que sus alumnos aprendan.			
2.3 Utiliza la evaluación de los aprendizajes con fines de mejora.			
2.4 Construye ambientes favorables para el aprendizaje.			
<b>Dimensión 3. Un docente que se reconoce como profesional que mejora continuamente para apoyar a los alumnos en su aprendizaje</b>	3.1.3,3.2.2,3.3.1,3.3.2, 3.3.3,3.3.4	3.1.1,3.1.2,3.2.3	3.2.1
3.1 Reflexiona sistemáticamente sobre su práctica docente como medio para mejorarla.			
3.2 Emplea estrategias de estudio y aprendizaje para su desarrollo profesional.			
3.3 Utiliza diferentes medios para enriquecer su desarrollo profesional.			
<b>Dimensión 4. Un docente que asume las responsabilidades legales y éticas inherentes a su profesión para el bienestar de los alumnos</b>	4.1.2, 4.1.3, 4.2.1, 4.2.2, 4.2.3, 4.2.4, 4.2.5.	4.1.1, 4.4.1.	4.3.1, 4.3.2, 4.3.3, 4.3.4.
4.1 Considera los principios filosóficos, los fundamentos legales y las finalidades de la educación pública mexicana en el ejercicio de su función docente.			
4.2 Establece un ambiente favorable para la sana convivencia y la inclusión educativa en su práctica docente.			
4.3 Considera la integridad y seguridad de los alumnos en el aula y en la escuela.			
4.4 Demuestra altas expectativas sobre el aprendizaje de todos sus alumnos.			
<b>Dimensión 5. Un docente que participa en el funcionamiento eficaz de la escuela y fomenta su vínculo con la comunidad para asegurar que todos los alumnos concluyan con éxito su escolaridad</b>	5.1.1, 5.1.2, 5.1.3, 5.1.4, 5.2.1, 5.2.2, 5.3.1, 5.3.2.	No se aborda en el instrumento	No se aborda en el instrumento
5.1 Realiza acciones en la gestión escolar para contribuir a la calidad de los resultados educativos.			
5.2 Propicia la colaboración de los padres de familia y de distintas instituciones para apoyar la tarea educativa de la escuela.			
5.3 Considera las características culturales y lingüísticas de la comunidad en el trabajo de la escuela.			

\* Esta tabla sirve de referencia para identificar los aspectos a evaluar de los docentes de Preescolar, Secundaria, Telesecundaria, Educación Física, Segunda Lengua: Inglés, Asignatura Francés, Tecnología, Artes: Artes Visuales, Música, Teatro y Danza.

Fonte: SEP 2017a, p. 9

## 14. Modelos de Auto eficacia: escala de Hoy e Woolfork de Eficacia do Professor

Por favor indique sua opinião pessoal sobre cada afirmação circulando a resposta apropriada à direita	1 indica concordância forte, 2 concordância moderada até 6 que é desacordo forte
1. O quanto que um estudante pode aprender é principalmente relacionado ao seu background familiar	1 2 3 4 5 6
2. Se os estudantes não são disciplinados em casa é provável que não aceitem nenhuma disciplina	1 2 3 4 5 6
3. Quando eu realmente tento, eu posso passar pelos estudantes mais difíceis	1 2 3 4 5 6
4. Um professor é muito limitado no que pode realizar porque o ambiente de casa dos estudantes é uma forte influência no seu desempenho	1 2 3 4 5 6
5. Se os pais fizessem mais pelas suas crianças eu poderia fazer mais	1 2 3 4 5 6
6. Se um estudante não lembra da informação que eu dei na classe passada eu sei como aumentar a sua retenção na próxima lição	1 2 3 4 5 6
7. Se um estudante na minha classe se torna bagunceiro e barulhento, eu tenho confiança que conheço técnicas para redirecioná-lo rapidamente	1 2 3 4 5 6
8. Se um dos meus estudantes não tiver feito o tema eu sou capaz de avaliar com precisão se o tema estava no nível correto de dificuldade	1 2 3 4 5 6
9. Se eu relamente tentar eu posso chegar até mesmo aos estudantes mais difíceis e desmotivados	1 2 3 4 5 6
10. Quando chega no final um professor não pode fazer muito realmente porque a maior parte da motivação e desempenho de um estudante depende do que acontece dentro do seu ambiente familiar	1 2 3 4 5 6

# 15. INSTRUMENTOS DE AVALIAÇÃO DE PROFESSORES EM PORTUGAL

## ANEXO I

### Avaliação externa do desempenho docente

#### Guião de observação da dimensão científica e pedagógica

Agrupamento de Escolas/Escola: \_\_\_\_\_

Docente: \_\_\_\_\_; Grupo de recrutamento: \_\_\_\_\_

Observação n.º \_\_\_\_ Data: \_\_\_\_ / \_\_\_\_ / \_\_\_\_ Tema \_\_\_\_\_

Tendo em consideração as circunstâncias concretas de ensino e a especificação dos parâmetros de avaliação, os registos derivados da observação devem incidir nos comportamentos do docente avaliado.

Parâmetros	Especificação	Registos	
		Positivos	Negativos
Científico (Domínio)	Conteúdo(s) disciplinar(es)		
	Conhecimentos que enquadram e agilizam a aprendizagem do(s) conteúdo(s) disciplinar(es)		
Pedagógico (Segurança)	Aspectos didáticos que permitam estruturar a aula para tratar os conteúdos previstos nos documentos curriculares e alcançar os objectivos seleccionados, verificar a evolução da aprendizagem, orientando as actividades em função dessa verificação e acompanhar a prestação dos alunos e proporcionar-lhe informação sobre a sua evolução		
	Aspectos relacionais que permitam assegurar o funcionamento da aula com base em regras que acautelem a disciplina; envolver os alunos e proporcionar a sua participação nas actividades; estimulá-los a melhorar a aprendizagem		
Considerações:			



## ANEXO II

### Avaliação externa do desempenho docente

#### Classificação da observação de aulas

Agrupamento de Escolas/Escola: \_\_\_\_\_

Docente: \_\_\_\_\_; Grupo de recrutamento: \_\_\_\_\_

Parâmetros	Especificação e ponderação		Descrição	Classificação
Científico (50%)	Conteúdo(s) disciplinar(es)	40%		
	Conhecimentos que enquadram e agilizam a aprendizagem do(s) conteúdo(s) disciplinar(es)	10%		
Pedagógico (50%)	Aspectos didáticos	40%		
	Aspectos relacionais	10 %		
Apreciação global:				
Recomendações:				
O avaliador:			Classificação final (Escala: 1 a 10)	
-----/-----/-----			Nível:	

## ANEXO III

### Parâmetros científicos e pedagógicos e níveis de desempenho

Parâmetros	1. Científico	2. Pedagógico
	Níveis de Desempenho	<p><b>Tendo em conta</b></p> <p>1.1. Os conteúdos disciplinares - 40%.</p> <p>1.2. Conhecimentos que enquadram e agilizam a aprendizagem dos conteúdos disciplinares - 10%.</p> <p><b>Nota:</b> Caso 1.2. não se aplique o 1.1. passará a ter a ponderação de 50%.</p> <p><b>O professor avaliado revela:</b></p>
<b>Excelente</b> (9 a 10 valores)	Domínio pleno dos conteúdos disciplinares e de conhecimentos funcionais.	Segurança inequívoca tanto em termos relacionais como didáticos.
<b>Muito bom</b> (8 a 8,9 valores)	Muito bom domínio dos conteúdos disciplinares e de conhecimentos funcionais.	Muito boa segurança em termos relacionais e pedagógicos.
<b>Bom</b> (6,5 a 7,9 valores)	Bom domínio dos conteúdos disciplinares e de conhecimentos funcionais.	Boa segurança em termos relacionais e pedagógicos.
<b>Regular</b> (5 a 6,4 valores)	Domínio regular dos conteúdos disciplinares e de conhecimentos funcionais.	Segurança regular em termos relacionais e pedagógicos.
<b>Insuficiente</b> (1 a 4,9 valores)	Falhas graves evidentes nos conteúdos disciplinares e de conhecimentos funcionais.	Falhas graves evidentes em termos relacionais e pedagógicos.

