

---

## **ESTUDO COMPARATIVO: O QUE AVALIAM AS AVALIAÇÕES INTERNAS E EXTERNAS?**

*Eliezer Pedroso da ROCHA<sup>1</sup>*

**RESUMO:** O objetivo deste artigo é mostrar o resultado do cotejamento feito entre a média final das notas individuais, apontadas por professores de língua portuguesa e de matemática em suas respectivas avaliações internas, de alunos da 3ª série, do ensino médio de escolas públicas estaduais, e a proficiência desses mesmos alunos aferida pelo Saesp, buscando elementos que indiquem a necessidade de um diálogo entre as duas avaliações, visando melhorar o desempenho do estudante. Foi feita uma análise, comparando os dois resultados a partir de três escolas. Foram encontrados elementos que apontam divergências e convergências, em função dessa classificação, dada a subjetividade intrínseca do processo avaliativo. Essas divergências e convergências estão no caráter das avaliações, principalmente, a interna. Daí a pergunta que dá título ao artigo: o que avaliam as avaliações internas e externas? O desempenho? O comportamento? As atitudes? Os resultados mostraram que um acompanhamento individual desse aluno, a partir dos resultados das avaliações externas, pode ser fundamental para ajudar a melhorar o seu desempenho e, conseqüentemente, o da escola como um todo.

**PALAVRAS-CHAVE:** Avaliação externa, avaliação interna, Saesp, Idesp.

### **A COMPARATIVE STUDY: WHAT DO THE INTERNAL AND EXTERNAL EVALUATIONS EVALUATE?**

**ABSTRACT:** The purpose of this article is to show the result of a comparison made between the final average of individual grades, identified by teachers of Portuguese

---

<sup>1</sup> Supervisor de ensino da rede estadual paulista. Correspondência: [Eliezer.rocha@educacao.sp.gov.br](mailto:Eliezer.rocha@educacao.sp.gov.br)

language and mathematics in their internal assessments, from students in the third grade, high school students from public schools, and proficiency of those students measured by Saresp, trying to gather evidence to suggest the need for a dialogue between the two evaluations in order to improve student's performance. An analysis comparing the two results from the three schools was made. Elements were found pointing differences and, due to this classification, given the inherent subjectivity of the evaluation process. These divergences and convergences are in the nature of evaluations, mainly in the internal ones. Hence, the question that gives title to the article: what do the internal and the external evaluations evaluate? The performance? The behavior? The attitudes? The results showed that an individual support for that student, from the results of the external evaluations, can be crucial to help improve his performance and, consequently, the school as a whole.

**KEYWORDS:** External evaluation, internal evaluation, Saresp, Idesp

## INTRODUÇÃO

*A avaliação educacional nunca é um todo acabado, auto-suficiente, mas uma das múltiplas possibilidades para explicar um fenômeno, analisar suas causas, estabelecer prováveis consequências e sugerir elementos para uma discussão posterior, acompanhada de tomada de decisões, que considerem as condições que geraram os fenômenos analisados criticamente (VIANNA, 2000, p. 18).*

Avaliação educacional é um tema complexo e, como tal, merece um olhar atento e constante dos órgãos competentes. O olhar somente da sala de aula parece não dar conta dessa complexidade. Desde sempre, os professores fazem avaliação dos seus alunos, mesmo que não tenham tido, em suas formações, uma disciplina específica que os ajudasse nessa questão. Só recentemente alguns cursos de licenciatura têm, em sua grade curricular, uma disciplina com o título de avaliação. A grande maioria dos professores faz avaliação com base em práticas de seus antigos professores ou em poucas leituras, que fazem sobre o tema. Como salienta Vianna (2000, p. 161), “a avaliação é um campo aberto à reflexão teórica e, ao mesmo tempo, que esse pensamento teórico é, necessariamente, a infra-estrutura que possibilita a prática da avaliação.”

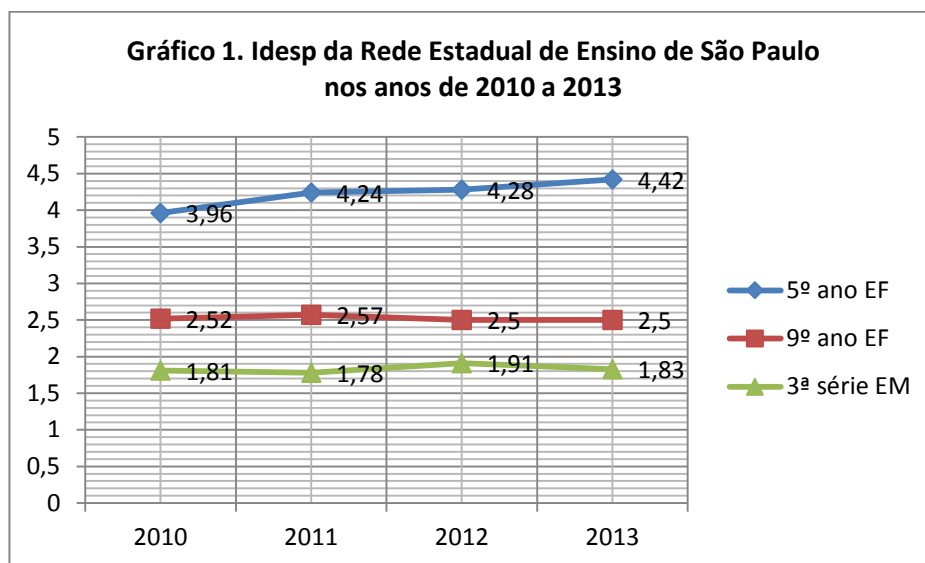
Em 2014, dando continuidade às discussões levantadas em 2010 e 2012, escolheu-se, para a 3ª edição do Conave – Congresso Nacional de Avaliação Educacional –, em Bauru, interior do Estado de São Paulo, o tema "**Avaliação de Sistema e Avaliação Escolar: aproximações necessárias**", apoiando-se na necessidade de promover um maior entendimento das peculiaridades das avaliações em larga escala e das desenvolvidas no interior das escolas. Como vemos, o diálogo entre as duas é necessário, não com o intuito de dizer que uma é melhor que outra ou dá melhores resultados, mas como complemento.

A educação é um dos temas prediletos dos candidatos a governo, tanto em nível federal, quanto estadual e até municipal, no caso das capitais e grandes municípios. Os governos procuram argumentos para explicar o porquê de sua rede de ensino estar no nível em que está, seja acima ou abaixo do esperado. A busca de argumentos dificilmente revela ações que retratam as responsabilidades dos gestores públicos.

Os resultados dos últimos quatro anos do Idesp – Índice de Desenvolvimento da Educação de São Paulo<sup>1</sup> – mostram aquilo que deve ser uma preocupação dos gestores dessa rede. Nesse período, houve avanço somente nos anos iniciais do ensino fundamental, e uma estagnação tanto nos anos finais quanto no ensino médio (gráfico 1). As metas, estabelecidas pela Secretaria da Educação do Estado de São Paulo para ser atingida no ano de 2030, são: 7,0 para o 5º ano do ensino fundamental, 6,0 para o 9º ano do ensino fundamental e 5,0 para a 3ª série do ensino médio. Pelo que apresenta o gráfico a seguir, não será nada fácil atingir esses índices tanto no 9º ano do ensino fundamental quanto na 3ª série do ensino médio, colocando em dúvida as políticas públicas desenvolvidas para esses segmentos. O que está sendo feito com os anos iniciais que poderia ser estendido aos demais ciclos? São as ações ou as provas estão direcionadas? São somente questionamentos, já que o objeto deste estudo é outro. Vejamos o gráfico:

---

<sup>1</sup> O Índice de Desenvolvimento da Educação de São Paulo (Idesp) é um dos principais indicadores da qualidade do ensino na rede estadual paulista. Criado em 2007, o índice estabelece metas que as escolas devem alcançar ano a ano. Os objetivos de cada escola são traçados, levando em consideração o desempenho dos alunos no Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo (Saresp) e o fluxo escolar de cada ciclo. Por esse motivo, a unidade escolar tem de oferecer uma meta diferente para cada ciclo. Disponível em <http://www.educacao.sp.gov.br/idesp>. Acesso em 25 out 2014.



Como o índice foi estabelecido em 2007, para ser atingido em 2030, e a última referência é de 2013, já foi percorrido pouco mais de 25% do total. Os anos iniciais (5º ano) do ensino fundamental já atingiram 63% da meta, dando indícios de que chegarão aos 100% no prazo estipulado. Já os anos finais (9º ano), apesar de ter atingido 42% no primeiro quarto de medição, têm os mesmos 42% verificados em 2010. Esses dados evidenciam a preocupação com a estagnação. Com o ensino médio, além de estar no mesmo patamar de 2010, teve uma ligeira queda de 2012 para 2013. Atingiu 37% da meta, com indícios de que não chegará ao que foi estabelecido para 2030. Pelos resultados, se continuarem as mesmas ações para esses dois segmentos da escolarização básica, a rede estadual ficará longe do que foi estipulado. Por esse motivo, faz-se necessário o desenvolvimento de ações, que deem continuidade, nos anos finais do ensino fundamental e no ensino médio, ao que é feito nos primeiros anos do ensino fundamental. As equipes que coordenam esses segmentos precisam de um trabalho articulado, com estudos mostrando os erros e acertos das políticas desenvolvidas.

## O SENTIDO DE AVALIAR

Quando o tema avaliação vem à tona, alguns questionamentos se fazem necessários: O que é avaliar? Como avaliar? Por que avaliar? Para que avaliar?

Em resposta à primeira pergunta, Tyler (1950, p. 69, *apud* LUKAS e SANTIAGO, 2009, p. 289) vale dizer que se trata de um “processo para determinar em que medida os

objetivos educacionais foram alcançados.” Ou seja, uma das primeiras tarefas do avaliador é definir quais são os objetivos que se quer alcançar com determinada atividade, com determinado conteúdo e, a partir de então, estabelecer um programa que será desenvolvido, visando o fim a ser atingido. Cronbach (1963, p. 672, *apud* LUKAS e SANTIAGO, 2009, p. 289) define a avaliação educacional como um “processo de coleta e utilização da informação para tomada de decisões.” Isto é, de tudo aquilo que o professor conseguiu coletar dos seus alunos – provas, trabalhos, atividades individuais e em grupo –, faz-se necessário buscar alternativas para que a decisão tomada seja a mais coerente com o que se quer. Não adianta fazer toda a coleta, obter as informações, e nada fazer com elas. Por último, nessa primeira análise, buscamos a definição de Scriven (1967; p. 139, *apud* LUKAS e SANTIAGO, 2009, p. 289), para quem a avaliação é um “processo pelo qual se determina o mérito ou valor de alguma coisa.” Para Scriven (*Apud* Ristoff, 2003, pp. 24-25),

*o processo avaliativo determina valor ao invés de estudá-lo, como um container que determina o conteúdo, como um copo que dá forma à água que contém. Em outras palavras, Scriven deixa explícita uma visão de avaliação cuja valoração ocorre não a partir do que é encontrado, mas a partir do que se constrói sobre o encontrado, com a participação decisiva de uma carga valorativa já a priori incorporada às estruturas conceituais do avaliador.*

Logicamente que há outras tantas definições da avaliação. No entanto, para o que queremos neste artigo, estas nos bastam. Mesmo assim, ainda vamos trazer uma abordagem de Patton (1977), para quem a definição não é o mais importante e nem o nome a ser dado; tudo isso, diz ele, pode ser discutido e negociado. Porém, o que não pode ser negociado é que avaliação esteja fundamentada em dados. Não podemos fazer qualquer análise apenas com intuição.

Luckesi (2010, p. 38) nos chama a atenção para a questão do juízo de valor do processo avaliativo. Devido ao seu caráter permeado de subjetividade,

*encontramos a possibilidade arbitrária do estabelecimento e da mudança de critérios de julgamento, a partir de determinados interesses. Por exemplo, pode-se reduzir o padrão de exigência, se se deseja facilitar a aprovação de alguém; ao contrário, pode-se elevar o padrão de exigência se se deseja reprovar alguém.*

O ser humano está constantemente avaliando e sendo avaliado. Isso pode ser feito de um modo formal ou informal. O diretor de uma empresa pode avaliar seus funcionários usando uma planilha de ações e comportamentos desejados ou repulsivos, ou simplesmente fazendo uso de uma observação à distância. Logicamente que uma avaliação formal pode trazer mais elementos para um julgamento e tomada de decisão. Qualquer que seja a área de trabalho, não é possível – ou pelo menos não deveria ser – tomar uma decisão sem que seja feita uma avaliação.

No caso específico da escola, também os professores fazem avaliação formal e informal. Enquanto que aquela é registrada e objetiva, esta é subjetiva e não registrada. As atividades desenvolvidas nas aulas possibilitam que o professor julgue o desempenho do aluno em acordo ou desacordo com os objetivos estabelecidos previamente. Essas atividades podem ser registradas ou não. O registro pode fazer com que o resultado da avaliação tenha um caráter objetivo. Mas isso não garante a precisão, se não houver critérios também objetivos para essa análise. A experiência em determinado assunto ou tema pode fazer com que o olhar esteja educado para certas abordagens iniciais. No entanto, que sejam abordagens iniciais e não o julgamento definitivo. O importante, salientam Fitzpatrick, Sanders e Worthen (2004), é que ambas se complementem. Não se trata de dizer que uma é melhor que outra, mas que concorrem para o mesmo fim. Análises sem esse cuidado podem influenciar políticas de responsabilização única dos professores pelos resultados dos seus alunos nas avaliações externas ou em larga escala, desconsiderando as estruturas nas quais as escolas estão inseridas.

Quando o assunto é o porquê da avaliação, faz-se necessário entender a distinção entre o porquê e o para quê.

Por quê avaliar?

Porque o sistema, a estrutura, o diretor, o professor etc. exigem que seja feita uma avaliação de algo, de um curso, de um programa.

Para quê avaliar?

Retomando as palavras de Cronbach, citado anteriormente, a avaliação é feita para que uma decisão seja tomada. Ao se determinar o valor ou o mérito do que está sendo avaliado (Scriven, 1967), faz-se necessário tomar uma decisão. Retomando Tyler (1950), em se tratando de escola, a avaliação é feita para que seja confrontado o resultado aferido com o objetivo educacional proposto ou estabelecido.

Há dois tipos básicos de avaliação, que tem em Scriven seu sistematizador: a somativa e a formativa. A formativa “está diretamente relacionada com a decisão de desenvolver um programa, sua modificação ou, ainda, sua revisão.[...] A avaliação do tipo somativa está ligada à decisão de continuar com o programa, encerrá-lo ou adotá-lo, entre outros aspectos”(VIANNA, 2000, pp. 86-87). Em sala de aula, é comum o desenvolvimento da avaliação do tipo somativa. Como salienta Luckesi (2010, p.34), “o ato de avaliar não serve como pausa para pensar a prática e retornar a ela; mas sim como um meio de julgar a prática e torná-la estratificada”.

### **AValiação INTERNA E EXTERNA: O DIÁLOGO**

Em resposta à pergunta da pesquisa “o que avaliam as avaliações internas e externas”, inicialmente vamos defini-las. “Quando nos referimos a avaliações internas temos em mente as que são realizadas pelas escolas. É evidente que a avaliação na escola é parte do processo formativo, constituindo o trinômio ensino-aprendizagem-avaliação, sob orientação do professor” (VIANNA, 2003, p. 11). Quanto a isso, não há muita dúvida, supomos. Já no tocante à avaliação externa, há controvérsias. Ficamos com a definição de Alavarse (2013). Para esse pesquisador, as avaliações externas são assim denominadas

*porque são definidas, organizadas e conduzidas por quem não se encontra no interior das escolas, [...]. Essas avaliações externas, tendo em conta sua abrangência, também são denominadas de avaliações em larga escala, ressaltando sua visibilidade e, em decorrência, sua face de política pública em educação.*

Apesar do seu caráter de instrumento utilizado para elaboração e monitoramento das políticas públicas, é possível trazer para o interior das escolas seus resultados e fazer uma análise mais pormenorizada. É isso que iremos mostrar com este estudo. O embate está em definir o que pode e o que não pode ser divulgado ou publicado. Há um tipo de divulgação que é pública, em relatórios impressos e *on-line*, enquanto que outro pode ficar restrito ao interior das escolas, para que seus gestores e professores possam analisar e buscar alternativas para o acompanhamento individualizado do aluno. Cada escola teria uma senha, com acesso exclusivo, feito pelo diretor, aos dados referentes aos alunos que



fizeram a prova. Esse modelo não é original, haja vista que a Prova São Paulo já disponibilizava essa senha aos diretores da Rede Municipal de Ensino. Não é necessário disponibilizar os itens das provas, mas tão somente as proficiências individuais, mesmo que essas tenham um erro de medida que dificulta a exatidão. Será importante mostrar que a proficiência não é um dado absoluto, mas uma probabilidade estatística, um traço latente. Os itens das provas não podem ser disponibilizados em sua totalidade, para que a Secretaria não os “perca” para avaliações futuras. Contudo, a proficiência individual é disponibilizada para a rede Sesi.

O nosso estudo terá, como instrumento de análise de avaliação externa, o Saresp, que tem seu foco em leitura, em Língua Portuguesa, e em resolução de problemas, em Matemática. Assim sendo, não se trata de analisar todos os aspectos da Língua Portuguesa, nem os da Matemática. Quando se diz que o aluno tem proficiência  $x$  em Língua Portuguesa, essa proficiência é em leitura. A redação, no Saresp, é feita por amostragem. O seu resultado não é computado na composição do Idesp. Daí o cuidado quando se faz o cotejamento entre a avaliação externa e a interna. O professor de Língua Portuguesa, quando atribui uma nota ou um conceito ao aluno, faz uma análise geral desse aluno, principalmente em leitura e escrita. No caso da Matemática, o seu estudo é muito mais do que resolução de problemas. Esse alerta precisa ser feito para que não sejam tiradas conclusões apressadas, sem as devidas considerações. As avaliações externas são feitas a partir de um recorte da matriz de referência, assim como o professor também faz sua avaliação a partir de um recorte estabelecido por ele do que os alunos devem saber em função do que foi desenvolvido em sala de aula. Isto posto, tanto a avaliação interna quanto a externa são feitas a partir de recortes de um todo. Não é possível avaliar tudo o que é desenvolvido nem pelos professores, nem pelos sistemas. Daí a importância de se fazer um bom recorte que contemple aquilo que pode distinguir melhor o que sabe e o que não sabe determinado aluno. A partir do momento em que foi identificado que o aluno não aprendeu aquilo que deveria ter aprendido, retoma-se a atividade com outra metodologia para que ela, a aprendizagem, ocorra, pois, como destaca Demo (2004, p. 74), “avalia-se para garantir o direito de aprender”.

Os gráficos abaixo mostram a classificação, por nível de proficiências, das disciplinas avaliadas pelo Saresp, que é o Sistema de Avaliação do Rendimento Escolar do Estado de São Paulo, e um dos indicadores de desempenho da escola que compõem o



Idesp. Outro indicador é o fluxo. Há estudos para que o nível socioeconômico também seja um indicador.

Gráfico 2. Níveis de Proficiência de Língua Portuguesa do Saresp				
Níveis de Proficiência	5º EF	7º EF	9º EF	3ª EM
Abaixo do Básico	< 150	< 175	< 200	< 250
Básico	150 a < 200	175 a < 225	200 a < 275	250 a < 300
Adequado	200 a < 250	225 a < 275	275 a < 325	300 a < 375
Avançado	≥ 250	≥ 275	≥ 325	≥ 375

Gráfico 3. Níveis de Proficiência de Matemática do Saresp				
Níveis de Proficiência	5º EF	7º EF	9º EF	3ª EM
Abaixo do Básico	< 175	< 200	< 225	< 275
Básico	175 a < 225	200 a < 250	225 a < 300	275 a < 325
Adequado	225 a < 275	250 a < 300	300 a < 350	325 a < 400
Avançado	≥ 275	≥ 300	≥ 350	≥ 400

De acordo com a descrição, o nível adequado é o esperado para alunos que estejam no final de cada segmento da escolarização básica. Mesmo que, no nível básico, ele tenha estrutura para interagir com o que será desenvolvido no ano/série seguinte, o domínio do conteúdo é mínimo.

No sumário executivo do Saresp 2013, da Secretaria da Educação do Estado de São Paulo (p. 4), os níveis de proficiência são assim descritos:

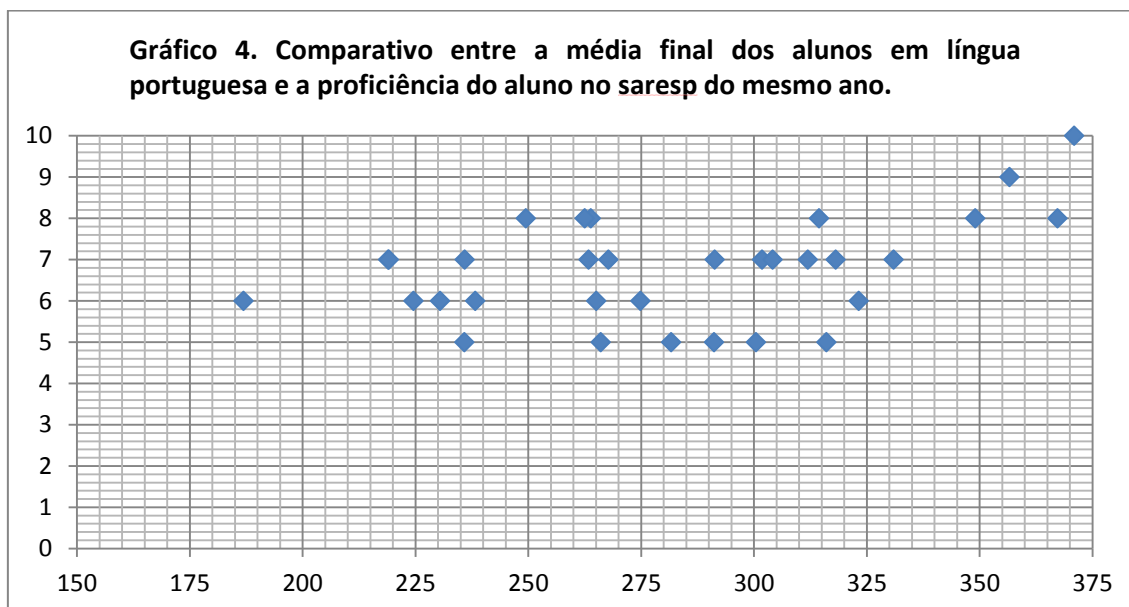
**Abaixo do Básico:** Os alunos demonstram domínio insuficiente dos conteúdos, das competências e habilidades desejáveis para o ano/série escolar em que se encontram.

**Básico:** Os alunos demonstram domínio mínimo dos conteúdos, das competências e habilidades, mas possuem as estruturas necessárias pra interagir com a proposta curricular no ano/série escolar subsequente.

**Adequado:** Os alunos demonstram domínio pleno dos conteúdos, das competências e habilidades desejáveis para o ano/série escolar em que se encontram.

**Avançado:** Os alunos demonstram conhecimentos e domínio dos conteúdos, das competências e habilidades acima do requerido para o ano/série escolar em que se encontram.

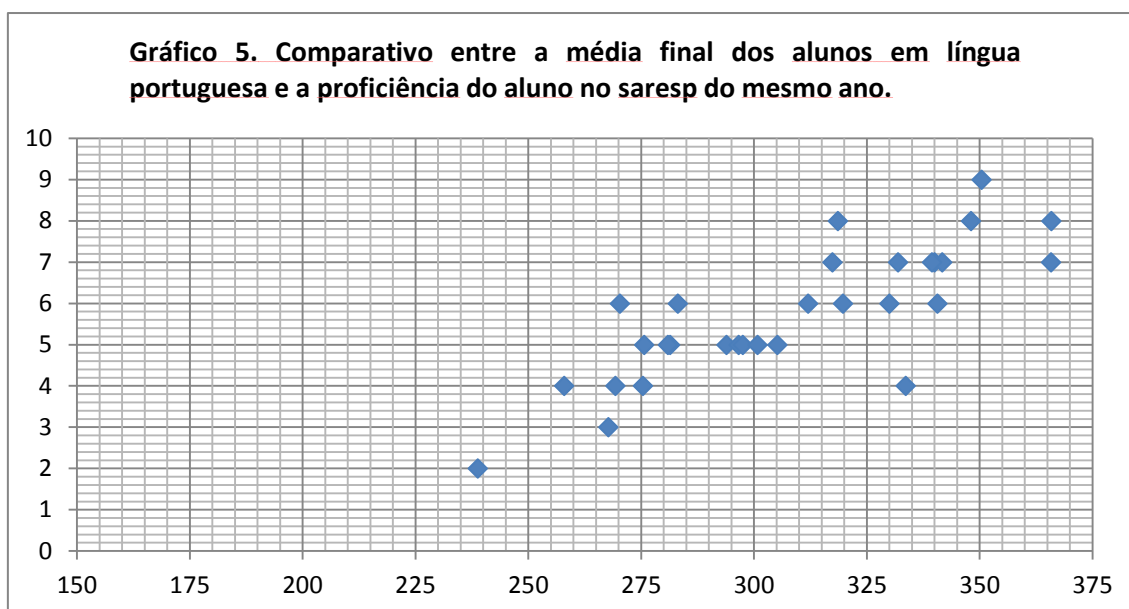
No gráfico que apresentamos a seguir, veremos um primeiro comparativo entre as duas avaliações. Apesar de termos feito a introdução abordando o Idesp, porque é o indicador que estabelece os índices anuais e as metas que as escolas têm de atingir até 2030, no decorrer das próximas páginas vamos nos referir especificamente ao Saresp.



Os primeiros gráficos são o resultado de estudos com três escolas. Pela análise dos gráficos 4 e 7 (3as. séries A e B), podemos perceber que não basta a escola ter acesso ao resultado por série, mas por turma, dada a diferença que se estabelece no interior dessas turmas, exigindo um trabalho diferenciado.

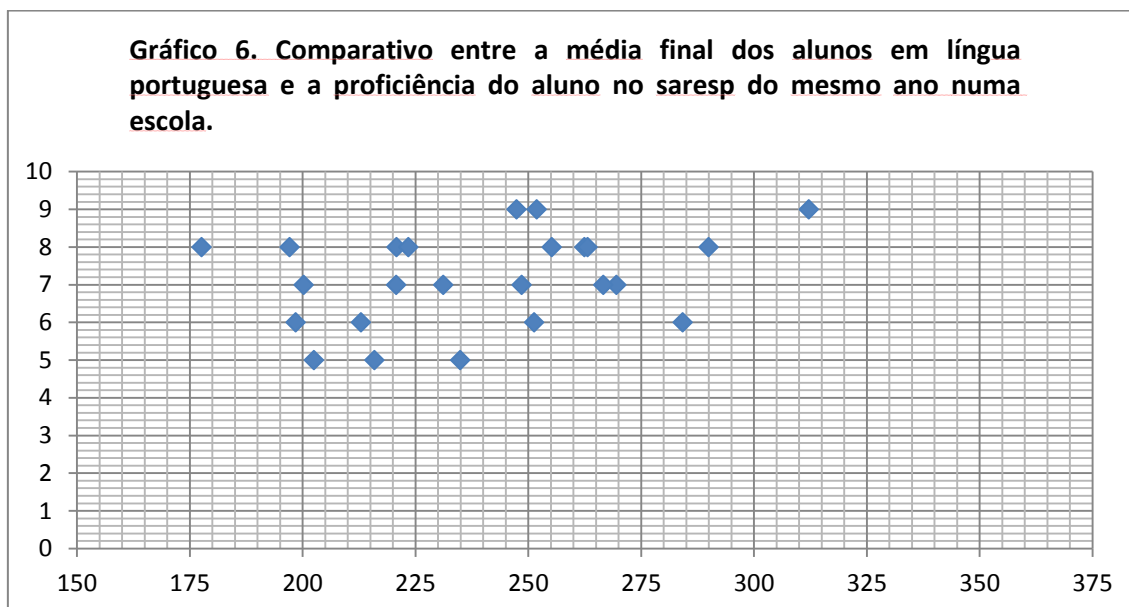
O gráfico 4 mostra que, em Língua Portuguesa, todos os alunos foram promovidos, pois não há alunos com média abaixo de cinco indicada pelo professor. Era uma sala de aula onde 31 alunos foram avaliados, tanto pelo professor quanto pelo Saresp. Desse total, somente um aluno ficou com média final 10,0 (dez). Esse aluno foi também o que apresentou o melhor desempenho no Saresp, com proficiência 370,9. Por outro lado, houve um aluno que teve média final 6,0 (seis), como resultado das avaliações internas realizadas pelo professor, com proficiência 186,9. Foi a menor proficiência aferida pelo Saresp para essa turma. Contudo, nessa mesma turma, tivemos um aluno com proficiência 316,5, que teve média final 5,0 (cinco). Essa proficiência significa que o aluno está no nível adequado para a série em Língua Portuguesa. Apesar de sua proficiência indicar que ele está aprendendo o que é adequado para a série, não é isso que a média indicada pelo professor demonstra. Uma pergunta que fica é: qual a razão desse

aluno ter 5,0 na média de suas avaliações internas e, na avaliação externa, ter uma proficiência considerada adequada para a turma e mais elevada que muitos dos seus colegas, que ficaram com média 6, 7 e 8? O resultado desse aluno internamente não é fruto apenas do desempenho ou da aprendizagem. Outros fatores parecem interferir na composição da média do professor. Que fatores? Só o professor pode dizer. Pelo gráfico é possível verificar que há uma coerência em relação às duas avaliações, quando os 4 alunos com maior média na avaliação interna também são os que têm maior proficiência. Todavia, há casos que necessitam de um estudo mais detalhado, pois médias elevadas como 8,0 (oito) foram atribuídas a alunos com proficiência inferior a 275, que é o adequado para alunos que estão no 7º ano do ensino fundamental.

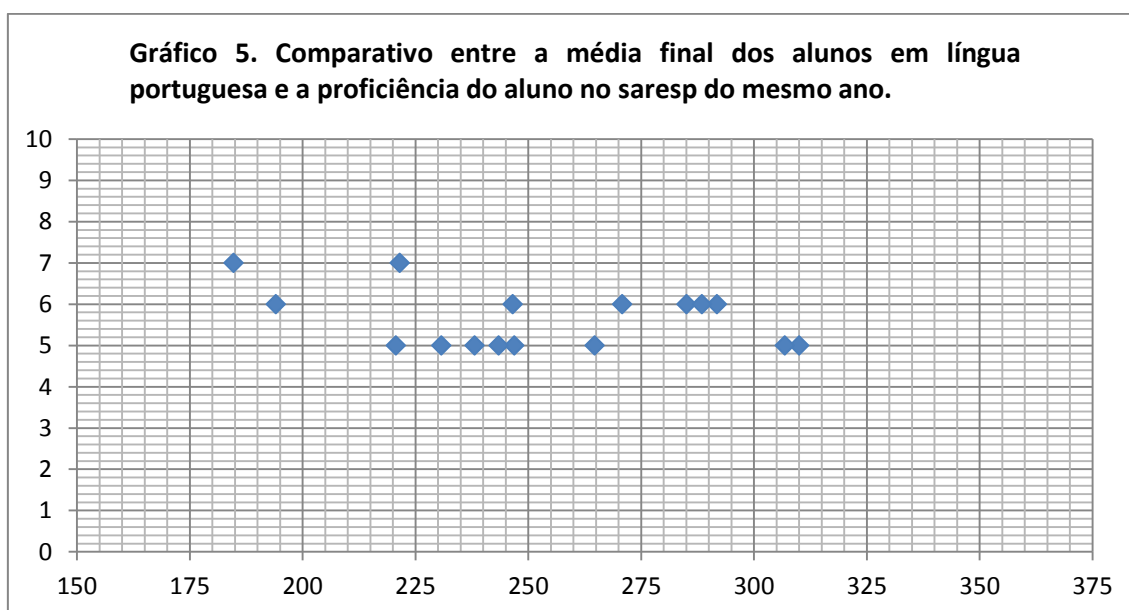


Nesse gráfico 5 é possível perceber certa coerência no todo entre as duas avaliações, pois o desenho mostra que os alunos com maior proficiência no Saesp são os que têm melhor desempenho na avaliação do professor. Nessa turma, nem todos os alunos foram promovidos. Temos seis alunos retidos no final do ano. Porém, como o Saesp foi feito no mês de novembro, esses alunos fizeram a prova. O que causa estranheza, contudo, é que um deles, inclusive, teve uma das maiores proficiências de toda a turma e, mesmo assim, ficou retido nessa série. Esse aluno teve proficiência 334 e ficou retido, enquanto que alunos com proficiência inferior a 285 foram promovidos. É preciso lembrar que o resultado final do professor é fruto de 4 bimestres de estudos, de provas, trabalhos, ou

seja, de várias atividades avaliativas, enquanto que o Saresp é realizado em apenas dois dias, sendo um apenas para a disciplina em questão. No caso de Língua Portuguesa, também é feita a redação, mas em caráter amostral. O seu resultado não compõe a proficiência do aluno, demonstrada nos gráficos deste estudo.



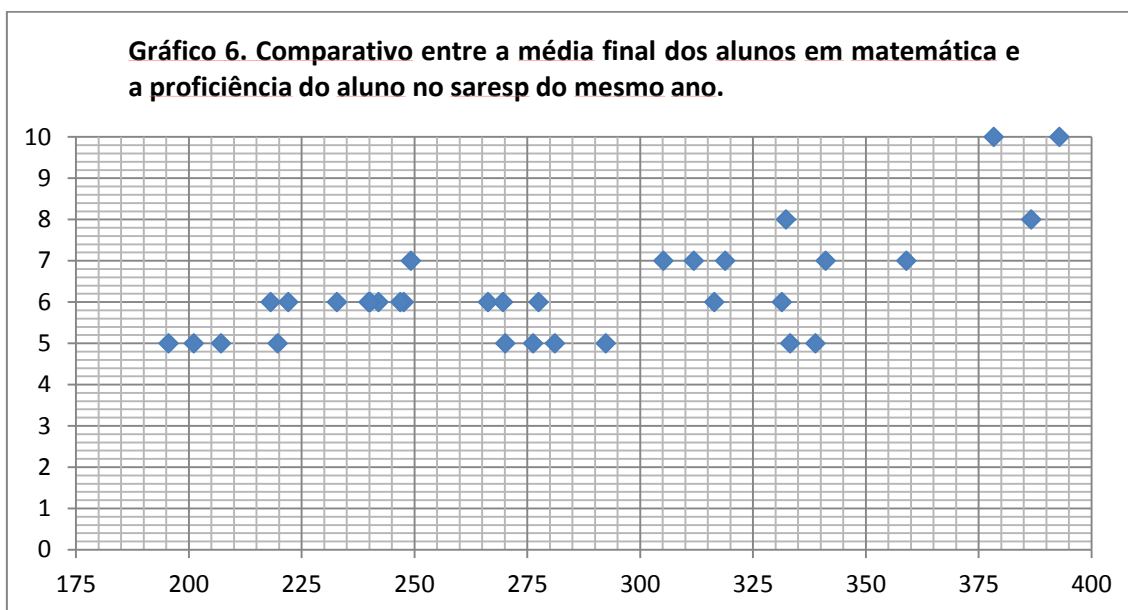
Nessa turma, como vemos pelo gráfico, todos os alunos também foram promovidos. Contudo, somente um aluno tem proficiência acima de 300, considerado adequado para a série em questão, em Língua Portuguesa. Dos 24 alunos da turma, dois tiveram proficiência aferida como adequada para o 9º ano do ensino fundamental, 11 para o 7º ano, e 10 para o 5º ano. Mesmo assim, como dito acima, nenhum aluno ficou retido.



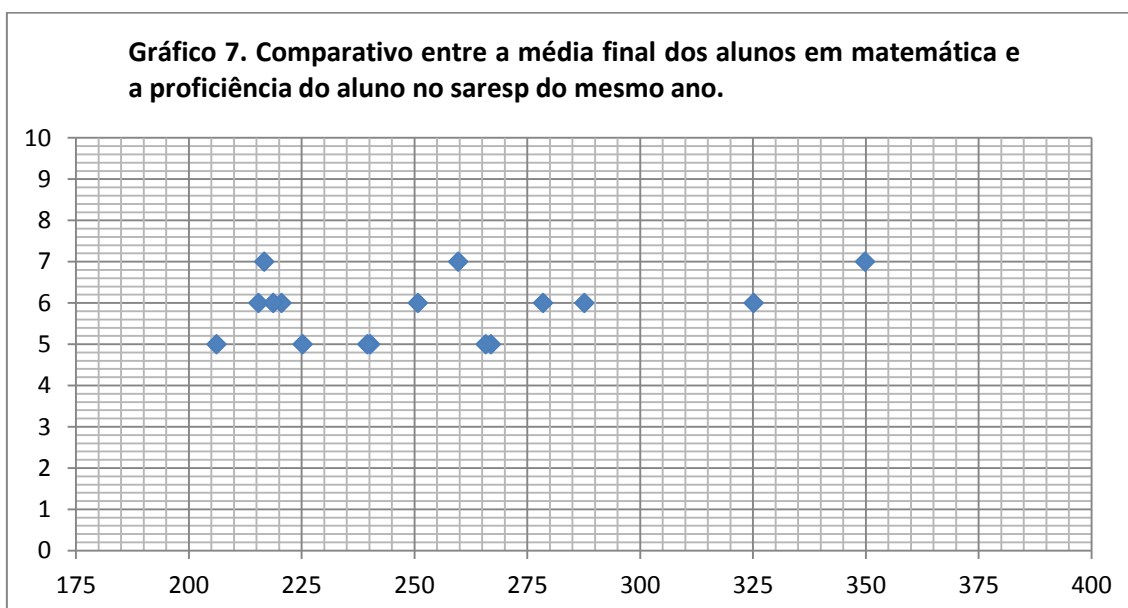
O gráfico 5 mostra que, nessa turma, aconteceu um fato interessante e curioso, que gera um olhar mais atento para o interior dessa escola. Todo professor deve ter os seus critérios de avaliação. No entanto, causa estranheza o fato de todos os alunos terem média mínima de 5,0 (cinco) e máxima de 7,0 (sete). Esse foi o limite estabelecido pelo professor. Voltando a Perrenoud (1999, p. 74), ele vai dizer que

*na mesma escola, professores acreditam firmemente que não se deve jamais dar a nota máxima, porque nenhum aluno alcança a excelência absoluta, ao passo que outros dão sistematicamente a melhor nota ao aluno que tem os melhores resultados, qualquer que seja seu nível absoluto. Pelas mesmas razões e devido à preocupação de não desconsiderar os alunos fracos, certos professores jamais recorrem às notas ou às apreciações mais desfavoráveis, enquanto que outros não hesitam em utilizá-las.*

As médias mais altas atribuídas pelo professor não foram aos alunos com proficiência mais elevada. Pelo contrário, os alunos com média mais alta são os que têm menor proficiência, enquanto que os que têm média mais baixa são os que têm maior proficiência. Um dos alunos, que ficou com a média final sete (7,0) com o professor, teve a menor proficiência no Saesp, ou seja, 185. Essa proficiência equivale ao nível básico para alunos do 5º ano do ensino fundamental. Apesar de o Idesp dessa escola ser alto, essa turma não contribuiu muito com esse resultado, pois somente dois alunos tiveram a proficiência aferida como adequada para a 3ª série do ensino médio. A grande maioria, sete alunos, dos 16 que foram avaliados, saiu com a proficiência adequada para o 7º ano do ensino fundamental. Isto mostra que fazer análise somente por série não é suficiente. Essa turma precisa de um trabalho diferenciado em relação aos seus colegas do período diurno. Essa escola teve a média elevada em função do desempenho da turma do período diurno.



O gráfico 6 é da mesma escola e turma do gráfico 2. Nele, podemos perceber que o professor de Matemática não teve alunos retidos em sua disciplina nessa turma. Os alunos com a média mais alta com o professor também foram os que tiveram a maior proficiência aferida pelo Saesp. Há uma concentração maior de alunos com nível de proficiência adequada para a série em questão. Mesmo assim, dois alunos que obtiveram a média mínima para serem promovidos estavam entre os com melhor desempenho no Saesp.



Da mesma forma que ocorreu com o professor de Língua Portuguesa, também o professor de Matemática dessa turma não teve aluno com média abaixo de cinco (5,0) e acima de sete (7,0) em sua disciplina. Causa estranheza que a mesma turma tenha o mesmo resultado tanto em Língua Portuguesa quanto em Matemática. Ainda que a maior proficiência aferida pelo Saresp tenha sido do aluno que teve 7,0 (sete), e a menor do aluno que teve 5,0 (cinco), podemos ver alunos com 6,0 (seis) e 7,0 (sete), mas com proficiência baixíssima para a 3ª série do ensino médio. Em Matemática, o adequado para a série está entre 350 e menor que 400. Esses alunos tiveram proficiência inferior a 275, considerada adequada para o 5º ano do ensino fundamental. Também em Matemática, houve pouca contribuição dessa turma para o Idesp da escola.

Pelo exposto, foi possível mostrar que esse cotejamento não só é possível, como, a nosso ver, necessário. A quem interessa que a proficiência individual não seja divulgada? Será que a divulgação pode colocar em xeque as avaliações, tanto externa quanto interna? Que tal avaliar a avaliação?

## **AVALIAR A AVALIAÇÃO**

*Avaliar o professor é sempre tarefa difícil e ingrata, mas deve ser feita, desde que com competência e, sobretudo, bom senso. A avaliação indireta, por meio do desempenho dos alunos, por sua vez, representa grande risco, com amplas consequências. (VIANNA, 2003, p. 11).*

Dentre as definições analisadas por Lukas e Santiago (2009, p. 289), vamos trazer, aqui, uma que apresenta um dado novo: a de Robert Stake. Para o pesquisador, a avaliação é “um processo sistemático de emissão de juízos baseados em uma descrição das discrepâncias entre as execuções de estudantes e professores e os padrões da aprendizagem e do ensino previamente definidos.” Como vemos, o dado novo é a avaliação também do ensino e não somente da aprendizagem, ou seja, avaliar o avaliador e não somente o avaliado. Em suma, avaliar também o professor e não somente o aluno.

Stake indica que também os professores têm um padrão previamente definido que deverá ser atingido. O ensino, diz ele, também deve ser avaliado. Quem define o que será avaliado ou qual é o patamar a ser atingido? No caso do professor, pode ser o sistema do qual ele faz parte. É nesse contexto que se insere a avaliação institucional, além,



logicamente, da avaliação externa da aprendizagem. Costumeiramente, avalia-se a aprendizagem, sem que se dê conta da necessidade da avaliação do ensino.

A prova, lembra-nos Moretto (2004), deve ser entendida como um momento privilegiado de estudo e não um acerto de contas. Em primeiro lugar, salienta Vianna (2003, p. 8), só se deve avaliar aquilo que foi efetivamente desenvolvido em sala de aula, haja vista a avaliação ser parte integrante do ensino e não somente da aprendizagem. Logicamente que uma avaliação pode ser feita de diversas formas e não apenas com a prova especificamente. Nada de provas surpresas para saber se o aluno estava prestando atenção na aula. Esse é um dos motivos de que, para avaliar, é necessária a coerência. Não se pode ficar no toma lá dá cá, fruto da visão bancária do ensino, como bem retratou Paulo Freire. A perspectiva construtivista defendida por alguns educadores mostra que o aluno não é mero receptor, mas construtor do conhecimento. Isto posto, o professor deve ter a clareza, na avaliação, de quais são os objetivos propostos para aquilo que está sendo avaliado. Quando vai aferir o conhecimento do aluno pelo instrumento da prova, deve ter parâmetros para tal, tem de ter critérios. Muitas vezes, um aluno não responde a uma questão da prova não porque não sabe o conteúdo, mas porque não entendeu o enunciado. Uma questão mal elaborada dificulta a resposta. Como não é comum fazer meta-avaliação, ou seja, avaliar a avaliação, essas questões são tidas como boas e o resultado nem sempre condiz com a expectativa do professor.

Anteriormente, vimos que, na descrição dos níveis de proficiência, as palavras-chave são conteúdos, competências e habilidades. Os alunos, para serem promovidos ao ano/série seguinte, devem ter o domínio desses elementos. Porém, questiona Vianna (2003, p. 8), será que nosso sistema educacional dá conta do desenvolvimento de habilidades e competências? Será que os professores foram ou estão preparados adequadamente para a identificação das habilidades e competências desenvolvidas por seus alunos?

No capítulo que trata da meta-avaliação, Fitzpatrick, Sanders e Worthen (2004) iniciam dizendo que toda avaliação tem um viés de princípio. Em primeiro lugar, porque são os avaliadores que definem qual será o método ou o instrumento que será usado para a avaliação. É a partir dessa definição que todo o resto é decidido. A importância da meta-avaliação (FITZPATRICK, SANDERS e WORTHEN, 2004, p. 442) está em, no caso da formativa, prover o avaliador de estudos prévios, melhorando o seu processo antes que

seja tarde demais. No caso da somativa, a meta-avaliação serve para adicionar credibilidade ao resultado final. Nos dois casos, se bem feita, serve para aumentar o potencial de qualquer avaliação.

Os estudos acerca da meta-avaliação formal começam nos anos 1960. Tanto os procedimentos quanto os critérios são colocados à prova, com o intuito de distinguir as boas das más avaliações. Apesar desse tempo, ainda não é comum ouvir que os professores colocam suas provas à prova para que outros professores façam suas análises e deem um *feedback* das mesmas. Em se tratando das avaliações externas, essa meta-avaliação poderia ser feita por uma auditoria, também externa ao processo, com especialistas no assunto. Caso não seja feita por auditoria, os próprios sistemas deveriam estar providos de profissionais capacitados que poderiam/deveriam fazer esse estudo. Como salientado acima, o objetivo da meta-avaliação não é dizer que está tudo errado, mas aumentar o potencial da avaliação. Quando os sistemas têm seus próprios meta-avaliadores e com competência para tal, o resultado final da avaliação passa a ter maior credibilidade. No entanto, para o público em geral, quando os meta-avaliadores são externos ao processo, isso mostra seriedade e compromisso com a melhoria do sistema. Quando se avalia o próprio trabalho, o viés é mais evidente ou torna-se mais difícil se livrar da pessoalidade e parcialidade.

## **CONSIDERAÇÕES FINAIS**

Abordar o tema avaliação é se colocar num ambiente extremamente complexo e desafiador. A partir da década de 1990, a avaliação externa passou a fazer parte de estudos de pesquisadores nacionais, fundamentalmente, nos institutos e fundações de pesquisa. As salas de aula da educação básica passavam ao largo dessas discussões. A Conferência realizada em Jomtiem, na Tailândia, trouxe à baila a necessidade de fazer com que a educação saísse dos discursos governamentais e passasse a fazer parte dos programas de governo, com o intuito de oferecer, de fato, educação para todos.

Contudo, educação para todos com a devida qualidade para que os estudantes pudessem participar da vida social e cultural da humanidade como um todo. Não bastava colocar os alunos em sala de aula. Se fazia necessário que o ingresso fosse acompanhado da aprendizagem efetiva.

No entanto, como aferir a aprendizagem efetiva? Como saber se os alunos estavam aprendendo, de fato?

É nesse contexto que as avaliações externas vêm contribuir com esse questionamento. Quando essas avaliações começaram a fazer uso da metodologia denominada Teoria da Resposta ao Item, foi possível traçar um acompanhamento longitudinal dos resultados do desempenho dos alunos, além de organizar uma série histórica.

Essa metodologia possibilita a colocação dos itens das provas numa escala de proficiência. Esses itens foram previamente testados. É a busca de maior objetividade nos resultados.

A disseminação dessas avaliações externas trouxe um elemento novo, até então pouco discutido no interior das escolas: como está sendo feita a avaliação do aluno pelo professor? É possível comparar a avaliação feita pelo professor com os resultados do Saresp, por exemplo?

Nosso trabalho buscou, a partir de três escolas, fazer uma reflexão acerca da possibilidade de comparação entre esses dois tipos de avaliação, não para endeusar uma e satanizar outra, mas para mostrar que esse diálogo é necessário e pode ajudar a melhorar o desempenho do aluno, ou, ainda, fazer com que o aluno aprenda, que é a principal razão do processo de escolarização.

Os resultados mostraram que o acompanhamento feito a partir do desempenho demonstrado no Saresp pode ajudar o professor e a gestão escolar a traçar alternativas de trabalho com o aluno e/ou com a turma. No entanto, mais estudos devem ser feitos que podem identificar as alternativas a serem realizadas e desenvolvidas no interior das escolas.

## REFERÊNCIAS

ALAVARSE, Ocimar Munhoz. Avaliar as avaliações em larga escala: desafios políticos. **Educação on line**, São Paulo, 12 dez. 2013. Disponível em <http://revistaeducacao.uol.com.br/textos/0/avaliar-as-avaliacoes-em-larga-escala-desafios-politicos-302490-1.asp>. Acesso em 26 out. 2014.

DEMO, Pedro. **Ser professor é cuidar que o aluno aprenda**. 2ª ed. Porto Alegre: Mediação, 2004.

FITZPATRICK, Jody L.; SANDERS, James R. & WORTHEN, Blaine R. **Program evaluation: alternative approaches and practical guidelines**. 3rd ed. Boston: Pearson Education, 2004.

LUCKESI, Cipriano Carlos. **Avaliação da aprendizagem escolar**. 21ª ed. São Paulo: Cortez, 2010.

MORETTO, Vasco Pedro. **Prova: um momento privilegiado de estudo – não um acerto de contas**. 4ª ed. Rio de Janeiro: DP&A, 2004.

PERREENOUD, Philippe. **Avaliação: da excelência à regulação das aprendizagens – Entre Duas Lógicas**. Trad. Patrícia Chittoni Ramos. Porto Alegre: Artmed, 1999.

RISTOFF, Dilvo I. Definições de avaliação. **Avaliação: Revista da Avaliação da Educação Superior**. v. 8, n. 1, mar. 2003, p. 24/25. Disponível em <http://educa.fcc.org.br/pdf/aval/v08n02/v08n02a03.pdf>. Acesso em 28 Out. 2014.

SÃO PAULO (Estado). Secretaria da Educação. **Sumário executivo**. São Paulo: Vunesp: FDE, 2013.

VIANNA, Heraldo Marelim. **Avaliação educacional: teoria – planejamento – modelos**. São Paulo: IBRASA, 2000.

\_\_\_\_\_. **Avaliações nacionais em larga escala: análise e propostas**. São Paulo: DPE, 2003.