

“A avaliação em larga escala no Brasil atingiu um platô”



Foto: Divulgação

Solange Monteiro, do Rio de Janeiro

Palco de políticas que transformaram o ensino público, como o caso de Sobral, o Ceará também é sede de um programa que visa tornar as avaliações educacionais mais detalhadas e frequentes, e que desde 2020 já é adotado no ensino médio do estado. O matemático Jorge Lira, coordenador do centro responsável por esse projeto, conta à *Conjuntura Econômica* como se deu o processo de elaboração dessas avaliações, os resultados obtidos em meio à pandemia, bem como o desdobramento dessa política em um

projeto piloto voltado aos anos finais do ensino fundamental, em que a inteligência artificial ajuda os professores a identificar déficits de conhecimento e habilidades dos alunos. Lira diz que a base para essa mudança em relação às avaliações é a elaboração de um bom currículo, feita junto com os professores. “Sem clareza sobre o que se pergunta, não se pode definir padrões de desempenho”, afirma, defendendo mais atenção ao letramento pedagógico- matemático dos professores, algo ainda pouco abordado no Brasil.

Conjuntura Econômica — Como surgiu o programa de avaliação formativa que hoje é aplicado no ensino médio da rede pública do Ceará?

Em um contexto mais amplo, tudo começou com uma iniciativa do estado do Ceará denominada Programa Cientista-Chefe, instituído em 2018, idealizado por Tarcísio Pequeno (*então presidente da Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico – Funcap*) e pelo economista Flávio Ataliba (*coordenador do Centro de Estudos para Desenvolvimento do Nordeste do FGV IBRE, ex-diretor-geral do Instituto de Pesquisas e Estratégia Econômica do Ceará – Ipece, e ex-secretário executivo de Planejamento desse estado*). O objetivo do programa nada mais é do que colocar a academia ao lado da gestão pública, fomentando a interação entre pesquisadores e gestores públicos para além de consultorias técnicas ou pesquisas puramente acadêmicas. Ou seja, não se trata apenas da entrega do resultado de uma pesquisa com começo, meio e fim definidos, mas do desenvolvimento de agendas comuns, atendendo a demandas de secretarias, formulando novas políticas públicas e buscando soluções conjuntamente. Da forma como foi posto, o programa começou a trabalhar com setores-chave como recursos hídricos, educação e saúde, com a ideia de também integrar esses aspectos da gestão numa base de dados comum. Então, além de haver políticas públicas formuladas e acompanhadas por pesquisadores e gestores, também é objetivo do programa criar e manter um fluxo de dados articulado numa espécie de *central big data* da gestão pública do Ceará.

Por conta dessa interação, começamos a atuar junto à Secretaria de

Educação (Seduc) em três pautas estruturantes: currículo, informação profissional e avaliação dos estudantes. Esses são grandes eixos do sistema educacional, a cargo de coordenadorias específicas, com cronogramas e atribuições fixas. Nosso trabalho era criar uma triangulação, buscando inovar em cada área a partir desse fluxo de informações, dados e evidências. Com isso, ajudamos a desenvolver um sistema de avaliação processual de aprendizagens. É sabido que tanto o governo federal quanto os estados promovem avaliações em larga escala – como o Saeb, cujos resultados foram recentemente divulgados –, mas essas são avaliações de fim de ciclo. Ou seja, são relatórios sobre o passado, um retrato do que foi a culminância de uma etapa de aprendizagem, mas que não contam todo o “filme” do processo. Então, a partir de 2018, iniciamos a cooperação com a Seduc, trazendo pessoas tanto de perfil pedagógico quanto especialistas em áreas quantitativas, como cientistas de dados, psicometristas e econometristas, para desenvolver um sistema dentro da rede pública que complementasse, de várias formas, a avaliação externa, ainda dependente de contratos dispendiosos e que trazem informação muito agregada e abreviada para o uso efetivo em sala de aula.

Foi nesse sentido que começamos, ao mesmo tempo, a minerar dados, tanto das avaliações externas que já existiam quanto os provenientes desse novo sistema, para cumprir nosso segundo objetivo, que era o de entender pedagogicamente essa informação quantitativa em alinhamento ao currículo. Em resumo, precisávamos transformar dados em evidências que o professor lesse, entendesse e apro-

priasse no seu entendimento sobre o que é o currículo, quais são os objetivos e a progressão esperada da aprendizagem. Isso é algo que o Brasil estranhamente não tem destacado: temos uma base curricular comum, mas sem especificações dos objetivos progressivos de aprendizagem. Esse trabalho foi, inicialmente, realizado com a ajuda de um grupo de 300 professores do ensino médio de toda a rede estadual, o que representou cerca de 10% da força de trabalho, com os quais começamos a atuar em polos regionais. Os professores traziam contribuições e levavam recomendações de materiais e sequências didáticas para testar em sala de aula. Com isso, fomos desenhando o currículo e o sistema de avaliação com a validação desses docentes, e em cerca de um ano e meio construímos o documento curricular e a matriz de avaliação integrados.

Mencionei, portanto, a avaliação e o currículo. A terceira frente do programa é a própria formação de professores, que aconteceu numa escala inicial desses 300 professores que cursaram uma pós-graduação *lato sensu* e, posteriormente, *stricto sensu*, tornando-se lideranças do processo de propagação desse sistema. O trabalho foi liderado pela Universidade Federal do Ceará junto a outras universidades públicas, federais e estaduais. Posteriormente, envolvemos nessa formação a Sociedade Brasileira de Matemática, que coordena o PROFMAT, mestrado nacional de matemática para professores das redes de ensino básico. Vários dos professores egressos passaram a ser líderes de formação de seus pares e, no início de 2020, a secretaria de estado encampou essa estrutura para dentro da política pública de formação de professores em serviço.

Quais elementos foram mais desafios nesse processo de formação?

Como mencionei, trabalhamos em uma especificação do currículo a partir de uma decisão pactuada com professores do que seria esperado em cada etapa, o que chamamos de mapa de progresso. Países como o Chile já fizeram esse dever de casa há pelo menos uma década, ao passo que ainda não temos isso. Do outro lado, elaboramos uma matriz especificando como fazer boas perguntas sobre o que se deve conhecer em cada etapa do aprendizado. Mas ainda identificávamos nos professores muitas fragilidades de repertório matemático e de transposição desse repertório para um contexto pedagógico. Por exemplo, ao fazer uma boa pergunta de matemática a um estudante, o professor normalmente usará o repertório de que dispõe para avaliá-lo, e uma abordagem que não necessariamente corresponde à meta de aprendizagem correspondente a esse problema ou exercício. O professor pode, por exemplo, adotar um critério mais focado no procedimento, deixando de investigar se o estudante teve uma compreensão conceitual profunda do que está por trás da utilização desse procedimento. Da mesma forma, se a tarefa e a evidência não forem devidamente usadas, o professor deixará de verificar se o estudante consegue mobilizar conhecimentos para modelar um dado problema, seja ele econômico, físico, sociológico, em termos da linguagem matemática. Essa seria uma das competências matemáticas realmente desejáveis.

O que tentamos buscar à época foi fortalecer o chamado conhecimento pedagógico do conteúdo – conceito sobre o qual fomos aprendendo mais

no processo, e que nada mais é do que a habilidade de conhecer bem o assunto e saber transmiti-lo bem. Há uma literatura enorme sobre conhecimento pedagógico do conteúdo que vem sendo produzida desde os anos 1980, mas que, estranhamente, não é foco das nossas licenciaturas, embora países com bons resultados na formação de professores trabalhem isso. Há pesquisas da OCDE que apontam o conhecimento pedagógico do conteúdo de professores como um dos fatores mais diretos e evidentes sobre a aprendiza-

“A partir de 2018, passamos a desenvolver um sistema dentro da rede pública que complementasse a avaliação externa, que traz informação muito agregada e abreviada”

gem, o que torna mais surpreendente o fato de ainda não tínhamos nada sistematizado sobre essa modalidade de conhecimento pedagógico-matemático em formações profissionais.

A que atribui o atraso brasileiro nesse campo?

Considero que seja um pouco de isolamento intelectual. Em áreas

como matemática ou economia, o Brasil se abriu aos temas e métodos com que o mundo trabalha. Nossa matemática, como pesquisa, se posiciona no panteão mundial em vários institutos de pesquisa e universidades. Isso se deve em parte à nossa abertura a aprender sobre o que é feito nas melhores instituições e publicações mundiais, o que permitiu criar uma matemática nacional robusta, mas com padrões internacionais. Na educação, no entanto, criou-se o mito da extrema especificidade da educação brasileira, que se firmou nos anos 1970 e pouco foi questionado desde então, em torno de chavões como “educação bancária”, os quais foram perpetuados em um debate insular em torno de um construtivismo algo ingênuo. Por exemplo, na formação de professores, ainda há resistência às práticas de mentoria e supervisão do trabalho docente, por conta de um conceito inadequado de autonomia do professor. Trata-se de discursos com ranços ideológicos e lastro científico mínimo, não amparado em evidências. Em alguns fóruns, por outro lado, a evidência quantitativa é rotulada como “positivista” e seu uso, de “gerencialismo”. Esse estado de coisas mental acaba por apreender o professor em formação numa teia de referências acadêmicas estranhas à realidade da sala de aula. Curiosamente, mesmo professores que passam por pós-graduações acabam, por vezes, mais distanciados de soluções para os problemas que enfrentam em seu cotidiano nas escolas.

No nosso caso, por termos matemáticos na equipe, contamos com as credenciais que a Sociedade Brasileira

de Matemática (SBM) e nosso Departamento de Matemática da Universidade Federal do Ceará foram criando ao longo do tempo. Veja o exemplo de ações educacionais como a Olimpíada Brasileira de Matemática das Escolas Públicas (Obmep). No momento de sua criação, a Obmep foi tida como algo que poderia instilar elitismo e mera competição na escola, desfavorecendo jovens em condições adversas, mas acabou se tornando um incrível motor de equidade educacional e social. O PROFMAT também é um exemplo dessas ações de ampla escala, pois se trata de um mestrado em ensino de matemática com componentes pedagógicos. Essas credenciais abriram um pouco a porta para nossas ações em redes públicas de ensino: o fato de ter matemáticos atuando em programas de educação básica já não era uma realidade tão estranha aos gestores dessas redes. Além disso, os professores começaram a ver sentido na nossa proposta porque fizeram parte de sua construção e a testaram nas salas de aula.

Como são as avaliações formativas?

Normalmente, avaliações de larga escala, como a do Saeb, são panorâmicas ou somativas, como se diz no jargão. Como um drone, sobrevoam a floresta e acompanham a cobertura vegetal. A avaliação processual na rede estadual, por sua vez, analisa também as “raízes”, ou seja, traz informações personalizadas de como a aprendizagem vai sendo desenvolvida ou, eventualmente, comprometida. Esse aprofundamento tornou-se possível por conta do currículo que foi previamente desenhado, dando clareza ao conteúdo que deve ser coberto

em cada etapa. São aplicadas três ou quatro avaliações ao ano, para uma base de 300 mil alunos. As escolas não são forçadas a fazê-la, mas a adesão é quase universal.

O teste pergunta sobre conhecimentos anteriores, para identificar lacunas pretéritas, e sobre as bases que serão fundantes para o que será trabalhado no período seguinte. Fazer uma avaliação em larga escala normalmente implica uma pergunta específica para avaliar um conjunto de conhecimentos encapsulado em

“Há pesquisas da
OCDE que apontam o
conhecimento pedagógico
do conteúdo de professores
como um dos fatores mais
diretos e evidentes sobre
a aprendizagem”

uma habilidade. Nos testes formativos, por sua vez, vamos investigando vários aspectos que compõem essa habilidade culminante e os processos cognitivos com que o estudante mobiliza os conhecimentos necessários: por exemplo, ler a informação que está expressa em um dado contexto, traduzi-la matematicamente, resgatar o repertório necessário, elaborar a estratégia de resolução do problema,

implementá-la e, por fim, fazer a análise sobre se aquilo que foi obtido nos procedimentos resolve o problema de forma plausível.

A estatística usada para isso parte de um modelo já usado no Saeb, no ENEM, a partir do qual utilizamos diversas variações. Nosso objetivo é o de abranger testes que evidenciem lacunas de aprendizagem – porque existem, são muitas e são graves –, mas que, por outro lado, representem níveis de complexidade cognitiva crescentes, a exemplo de testes como o PISA e o MARS. Buscamos ligar as duas pontas: seria mais cômodo, tecnicamente, apresentar apenas perguntas com baixas expectativas cognitivas. Embora desafiador, tentamos ampliar esse escopo, conduzindo estudantes e professores a tarefas mais representativas. Esse esforço não é vão: continuamos distantes dos níveis adequados no PISA, ao mesmo tempo em que percebemos que a avaliação em larga escala no Brasil atingiu um certo platô. Enquanto muitas redes ainda estão aquém das metas fixadas, municípios como Sobral já superaram essas metas há anos. No entanto, quando comparamos itens no “padrão PISA” a itens usualmente expostos em avaliações de larga escala, percebemos que os níveis mais elevados na escala dessas avaliações correspondem aos níveis iniciais de proficiência no teste internacional.

No sistema da rede estadual do Ceará, as avaliações são aplicadas tanto na forma impressa quanto *online*, dependendo da infraestrutura disponível. Hoje, mesmo com a maioria da informação sendo coletada aplicando testes impressos, em 15 a 20 dias os dados são entregues e analisados.

De que forma a falta de acesso a computadores e internet prejudica esse trabalho?

Hoje o percentual de avaliações feitas em formato digital é muito reduzido, algo que causa espécie se levarmos em conta que se trata apenas de que o estudante esteja com um computador ou *smartphone* conectado à internet. No Ceará, a pandemia levou a políticas emergenciais que derivaram, no estado e em alguns municípios, à entrega de *tablets*. Mas há várias vicissitudes nesse processo de avaliação em formato digital, desde a disponibilidade efetiva desses equipamentos nas atividades em sala de aula a questões de hábito ou de letramento digital para finalidades pedagógicas. Em muitas escolas no país, o aparato computacional e de conectividade ainda é muito precário.

Percebemos que a pauta das tecnologias educacionais enfatiza muito a ideia de “gamificação” e interatividade, muitas vezes sem intencionalidade pedagógica e avaliativa explícita e delimitada. O uso que fazemos de tecnologia e, em particular, de soluções de inteligência artificial (IA) tem por objetivo futuro o desenho de avaliações com esses aspectos interativos e lúdicos. No entanto, dado o atual estado de acesso limitado aos recursos digitais, temos usado IA mais expressamente no sentido de ajudar o professor a ter devolutivas rápidas e pedagogicamente ricas, de modo que o uso pedagógico das avaliações não seja obstruído caso a escola não esteja totalmente aparelhada e conectada à internet. Uma vez que tenhamos acesso e possamos interpretar os dados – e esse são aspectos em que a IA entra fortemente – conseguimos gerar devolutivas

quantitativas e qualitativas de forma muito mais localizada, tempestiva e precisa do que se o processo fosse completamente manual.

Vocês mantiveram essa política de avaliação durante a pandemia.

Quais resultados observaram?

As avaliações nos levaram a uma previsão, agora corroborada pelos resultados divulgados pelo INEP, de que a rede estadual conseguiria manter a média em matemática – um resulta-

“Na formação de
professores, ainda há
resistência a práticas de
mentoria e supervisão do
trabalho docente, por conta
de um conceito inadequado
de autonomia”

do desafiador, já que, diferentemente da língua portuguesa, a linguagem matemática, da forma como é abordada na escola, não é cotidianamente desenvolvida e empregada.

Essa manutenção da média foi confirmada na última avaliação do Saeb, em que as condições de aplicação foram mais próximas das de 2019. Mas há um aspecto ainda mais interessante, que a média não

revela. Quando abrimos os resultados para analisar a distribuição de proficiências, o que a série histórica de avaliações até 2019 nos mostra é que, entre os estudantes no primeiro quartil de proficiência, por exemplo, as lacunas mais evidentes eram em conhecimentos e habilidades muito básicas e que remontam à aritmética trabalhada nos anos iniciais do ensino fundamental. Nas últimas avaliações, apesar de a média ter sido mantida, houve uma mudança qualitativa. A dificuldade identificada hoje está mais situada em temas próprios no ensino médio do que era há quatro, seis anos. Ou seja, agora as dificuldades mais expressivas estão localizadas mais próximo de conhecimentos e processos cognitivos trabalhados nos anos finais do ensino fundamental e no ensino médio.

Isso evidencia que o trabalho cumulativo de recompor aprendizagens – ao qual já vínhamos nos dedicando, e que se tornou o alicerce das intervenções durante e após a pandemia – norteou ações pedagógicas em que os professores se sentem encorajados a voltar aos conhecimentos que os estudantes não consolidaram, mas sem perder de vista os objetivos curriculares. Houve, portanto, uma mudança qualitativa de resultados e de práticas.

De que forma o Centro de Excelência em Políticas Educacionais tem avançado no uso de inteligência artificial para aprimorar o programa de avaliação?

Com esse sistema criado e disseminado na rede estadual, fomos evoluindo para atender a outras demandas. Usualmente, os sistemas de uso mais

global, em que é preciso atender rapidamente a uma grande quantidade de estudantes, com devolutivas rápidas, os testes são, predominantemente, compostos de questões de múltipla escolha. Observamos as alternativas escolhidas pelo estudante e presumimos quais processos executou para chegar à resposta. De dois anos para cá, temos buscado outra direção, em projetos piloto voltados aos anos finais do ensino fundamental.

Em busca de uma evidência de aprendizagem mais robusta, apresentamos questões para o estudante em que ele tem que justificar as respostas que apresenta. O resultado, mais uma vez, é escrito em papel ou em meio digital. Obviamente, passar para o professor 10 a 15 mil dessas respostas em estado bruto representaria um trabalho colossal. Em vez disso, o que fazemos, juntamente com os professores, é identificar padrões de resposta com a ajuda de algoritmos de classificação. Normalmente, categorizamos respostas típicas de três padrões de desempenho, do insuficiente ao plenamente suficiente, com base na especificação de um objetivo de aprendizagem. Aqui vale reforçar o que mencionei anteriormente: sem clareza sobre o que se pergunta e o que se espera nas respostas, não se pode definir esses padrões de desempenho. Nossos cinco anos de trabalho pedagógico prévio foram fundamentais para embasar a definição de perguntas relevantes e de padrões de desempenho. Isso posto, a classificação é feita para um pequeno contingente de respostas que os professores categorizam e atribuem conceitos. Uma vez que se “ensina” a máquina sobre as respostas típi-

cas de cada categoria, o algoritmo vai nuclear o conjunto massivo de respostas em torno desses padrões, decidindo sobre quais são as respostas similares. Essa similaridade é interpretada cognitiva e pedagogicamente. Em suma, do oceano de respostas, surgem padrões condensados em três ou quatro grupos que contam com uma interpretação pedagogicamente definida, dada pelos próprios professores.

Com isso, o professor tem um mapa de calor de sua sala, um re-

“Temos usado IA para ajudar o professor a ter devolutivas rápidas e pedagogicamente ricas, de modo que o uso das avaliações não seja obstruído por falta de internet”

latório dinâmico e interativo, identificando as regiões “cognitivas” em sua turma que, por exemplo, demandam mais atenção quanto a certos conhecimentos e habilidades. Hoje esse projeto está sendo aplicado em redes municipais, entre as quais as das cidades de Sobral, Eusébio e Caucaia, em leitura e matemática, e já envolveram cerca de 20 mil alunos.

No Brasil, temos iniciativas bem-sucedidas no campo da educação pública, como a de Sobral, que se tornam referências inclusive em outros países. Mas não conseguimos disseminá-las em nível nacional. O que falta para que o país consiga dar escala a boas políticas de ensino?

Os projetos mencionados dependem de uma conjuntura política favorável presente no estado e nos municípios do Ceará e de outras redes com as quais temos interagido. Da mesma forma, dependem da abertura de grupos de pesquisa já extremamente consolidados na universidade a essas pautas de alta complexidade que, muitas vezes, não estavam, previamente, em suas agendas científicas. Por isso nos mobilizamos em criar o Centro de Excelência em Políticas Educacionais da Universidade Federal do Ceará em 2022, para que esse trabalho não ficasse dependente de uma conjuntura passageira, mas se perpetuasse para acolher e desenvolver políticas educacionais com a participação da academia, consolidando-se inclusive como uma interface e *hub* de pesquisa e inovação entre diferentes departamentos e conhecimentos. As universidades públicas tendem a ser departamentalizadas, e quisemos romper um pouco com esse desenho de escaninhos para atender a demandas reais e urgentes do setor público.

Ainda assim, manter um grupo coeso requer um programa em larga escala e que tenha continuidade. Nesse sentido, passamos a buscar parceiras. Temos conversado com a FGV DGPE e a EMAp para atuarmos conjuntamente, trazendo soluções que são escaláveis e sustentáveis, além de eficientes e de baixo custo por serem

baseadas em tecnologia muito mais acessível. A FGV DPGE, por exemplo, tem atuações bem-sucedidas em vários estados, e o prospecto comum é o de agregar nosso sistema de avaliação formativa e formação docente em seu repertório de soluções e sistemas de gestão educacional.

Em nossa trajetória recente, já temos acumulado parcerias, entre as quais com o Instituto Unibanco, com a Fundação Itaú Social, o Instituto Península (Movimento Profissão Docente), a Fundação B3 Social e o próprio MEC. Com elas, buscamos atuar em três frentes: na disseminação de ideias, para que ganhem em acreditação científica e implementação prática nas escolas; na sustentação financeira, para que as soluções cheguem com baixo impacto financeiro às redes públicas; e na consolidação das soluções com sua expansão e sustentação em escalas maiores. Claro que não pretendemos descrever uma receita de bolo, mas identificar elementos conceituais e processos estruturais.

Em nível federal, qual política pública poderia colaborar para a disseminação desse trabalho?

Diria que que uma das frentes prioritárias seria uma ação concertada para promover o letramento dos professores no conhecimento pedagógico do conteúdo no ensino de matemática. Além disso, o investimento federal e estadual em avaliação educacional poderia ser otimizado, aproximando a avaliação da instrução, ou seja, do trabalho docente. Isso, novamente, requer aprimorar a qualificação dos professores, o que demanda uma ação intensiva principalmente na formação para o ensino de matemática.

Há, nas redes públicas de ensino básico, muitos programas de formação de gestores e professores, ofertados por universidades ou por organizações de terceiro setor. Todavia, poucos dizem respeito ao exercício docente propriamente dito, ou seja, muitas iniciativas param na porta da sala de aula. Externamente, há ações importantes para uma gestão muito bem protocolada, com todos os fluxos desenhados e realizados, mas precisamos estar respeitosamente presentes na sala de aula, onde os professores

“Falamos da importância da educação integral para os estudantes, mas esquecemos que o professor deveria estar integralmente nessa educação, como um agente presente na escola”

precisam de apoio no enfrentamento de problemas de alta complexidade.

E quanto à inclusão digital?

Uma maior cobertura de acesso digital ajudaria a deslocar o esforço financeiro das redes para outras pautas, inclusive essa da formação profissional. Com isso, a logística de avaliações seria simplificada e te-

riamos devolutivas em horas em vez de meses. Recursos hoje retidos em esforço logístico colossal poderiam ser realocados para formar e fixar os professores em tempo integral nas escolas, no planejamento didático e no acompanhamento intensivo dos estudantes. Falamos da importância da educação integral para os estudantes, mas esquecemos que o professor deveria estar integralmente nessa educação, como um agente presente na escola.

Acrescento que, em nossa cultura, esperamos muito do governo. É urgente que a sociedade civil, na forma de instituições como a FGV, que estão estrategicamente na interface público-privada, tomem iniciativas e, eventualmente, a dianteira. É patente que o país está perdendo alguns bilhões por ano devido à falta de letramento matemático de nossos jovens. Um exemplo simples e anedótico disso: em um experimento avaliativo com educação financeira em uma rede estadual, perguntamos sobre cenários para minimizar perdas e maximizar ganhos em um negócio. Vimos que os estudantes sabiam calcular porcentagens em cálculos rotineiros de juros simples ou compostos, mas não conseguiam mobilizar esses conhecimentos para avaliar cenários de maior ou menor risco e tomar decisões com base nos cálculos que efetuam. Como empreender sem saber mensurar e comparar riscos? Isso é um problema no uso inteligente do repertório matemático.

Temos que nos imbuir desse dever de melhorar o ensino e a aprendizagem em matemática. Claro que o governo terá que aderir a isso, mas como sociedade, temos que tomar essa responsabilidade. ■