

Capítulo

1

Avaliação de Habilidades e Competências Baseada em Evidências e Jogos Digitais

Handerson Bezerra Medeiros, Eduardo Henrique da Silva Aranha e Isabel Dillmann Nunes

Abstract

The introduction of playful activities such as digital games with pedagogical purposes is a methodology to teach in a motivating way. Games stimulate learning and can evaluate the student without performing traditional tests. This approach, called stealth assessment, aims to obtain a correct and continuous assessment of players by discreetly monitoring students' knowledge, skills and attributes. In fact, games provide evidence about competency aspects of each action (choices, resolutions, and / or resolution attempts) made by the student. Stealth assessment uses an evaluation design framework called Evidence-Centred Design. In this context, this work aims to present to the public the use of games as a way to evaluate students, and the aspects of the stealth evaluations as a form of ongoing and cumulative assessment of student performance.

Resumo

A introdução de atividades lúdicas como jogos digitais com finalidades pedagógicas é uma metodologia para ensinar de forma motivadora. Os jogos estimulam a aprendizagem, podendo ser utilizados para avaliar o aluno sem a realização de testes tradicionais. Essa abordagem, chamada de stealth assessment (avaliação furtiva), tem como ponto principal obter uma avaliação correta e contínua dos jogadores, monitorando discretamente os conhecimentos, habilidades e atributos dos alunos. Stealth assessment utiliza uma estrutura de design de avaliação denominada Evidence-Centred Design. Neste contexto, o objetivo deste trabalho é apresentar o uso de jogos como forma de avaliar alunos, apresentando ao público os aspectos das avaliações stealth como uma forma de avaliação contínua e cumulativa do desempenho do aluno.

1.1. Introdução

Segundo Gatti (2005), a principal preocupação das escolas, pais e, sobretudo dos alunos, reside em descobrir quais as competências que os jovens precisam aprender hoje para alcançarem sucesso acadêmico, profissional e pessoal no futuro. Dados da Organização para Cooperação e Desenvolvimento Econômico (OCDE) encontraram que não somente as habilidades cognitivas, mas também as não cognitivas ou socioemocionais influenciavam fortemente as decisões educacionais [OCDE 2015].

A introdução de atividades lúdicas como jogos digitais com finalidades pedagógicas é uma metodologia para ensinar de forma motivadora, possibilitando uma prática significativa daquilo que está sendo aprendido e potencializando a construção e avaliação do conhecimento. Gatti (2005) afirma que o indivíduo se constitui e se desenvolve na medida em que interage e é estimulado pelo meio. Os jogos trazem em si uma infinidade de conceitos que estimulam a aprendizagem, pois permitem a interação e construção coletiva de conceitos e experiências, bem como a geração de conhecimento [Prensky 2012].

A aprendizagem baseada em jogos é considerada eficaz, pois utiliza técnicas de aprendizagem interativas (não provenientes de jogos) associadas a técnicas que já são utilizadas em jogos comerciais, como aprendizagem através do fazer, tentativa e erro, por descoberta, guiada, realização de tarefas, prática e *feedback*. Além disso, torna-se mais eficiente, pois está em conformidade com o estilo de aprendizagem atual dos alunos, sendo motivadora e versátil [Sena et al. 2016].

Avaliar a aprendizagem significa determinar as competências adquiridas por meio de aspectos quantitativos e qualitativos. A avaliação é um processo que envolve a coleta, medição e interpretação de dados. Com os valores coletados é realizada uma medição dos dados baseada em um conjunto de critérios, qualificando os resultados daqueles que foram avaliados. Interpretando tais resultados pode-se quantificar o grau de qualidade ou características que o discente possui; isso pode refletir diretamente na tomada de decisão dos educadores e determinar cada passo do processo ensino-aprendizagem.

Uma maneira de avaliar o aluno sem a realização de mais um teste é usar *stealth assessment* (avaliação furtiva) [Shute and Ventura 2013], a qual tem como foco as habilidades exigidas atualmente, como criatividade, colaboração e comunicação. O ponto principal desta proposta é obter uma avaliação correta e contínua dos jogadores, monitorando discretamente os conhecimentos, habilidades e atributos dos alunos.

A chave para a avaliação *stealth* é a sua natureza discreta aliada ao jogos. A ideia é que as escolhas e estratégias de um jogador estão constantemente e consistentemente informando o jogador de seu progresso e sucesso. Os jogos fornecem evidências sobre aspectos das competências com relação a cada ação (escolhas, resoluções e/ou tentativas de resoluções) executadas pelo aluno. Essas evidências possuem os valores que são coletados para serem analisados. *Stealth assessment* utiliza uma estrutura de design de avaliação denominada *Evidence-Centred Design* (ECD) [Almond et al. 2015].

O ECD define uma estrutura que consiste em três modelos: modelo de competências, de evidências e o de tarefas. Tais modelos funcionam em conjunto para coletar informações que permitem ao avaliador fazer inferências válidas sobre o que as

pessoas sabem, acreditam e podem fazer, e em que grau.

A vantagem do uso de ECD é que as avaliações são definidas conforme as competências que se deseja construir com os alunos, durante o processo de desenvolvimento dos jogos a serem utilizados. Desta forma, as tarefas disponibilizadas devem reverter em dados e informações que forneçam informações sobre as competências para essa avaliação. Além disso, a documentação gerada por esse design ajuda aos diferentes perfis que estão envolvidos na construção dos jogos aumentando a comunicação e a habilidade de desenvolver novas tarefas e avaliações paralelas às já existentes [Zieky 2014].

Assim, o objetivo deste capítulo é apresentar o uso de jogos como forma de avaliar alunos, mostrando os aspectos das avaliações *stealth* como uma forma de avaliação contínua e cumulativa do desempenho do aluno com o intuito de melhorar o processo de ensino aprendizagem com o auxílio do uso de informações geradas através da interação dos alunos com os jogos.

O material está organizado com o intuito de indicar que essas informações são úteis, pois indicam quais objetivos e/ou tarefas foram concluídas com sucesso e quais dificuldades encontradas, estabelecendo um *feedback* sobre o andamento do processo de ensino e aprendizagem. Apresentar os conceitos e arquitetura propostas do modelo ECD, bem como tecnologias para implementação de seus módulos, protocolos de comunicação e como são coletadas e geradas informações sobre o desenvolvimento do aluno com a execução desses módulos, permitindo que a prática docente se ajuste às necessidades discentes durante o processo de ensino e aprendizagem. Expor o uso de Redes Bayesianas para modelagem de competências e sua capacidade de inferência de conhecimento do aluno dada as informações geradas pelos jogos para cada competência. Por fim, a execução de uma dinâmica para implementar avaliações *stealth* desde a fase de concepção de tarefas, diagramação das competências usando as habilidades curriculares brasileiras até a análise dos resultados dos alunos.

1.2. Conhecimentos, competências e habilidades cognitivas e não cognitivas

As atuais diretrizes para educação, tal como Parâmetros Curriculares Nacionais¹, apontam para uma nova forma de ensino/aprendizagem, exigindo a mudança da tradicional transmissão do conhecimento e de memorização para um cenário com base nas competências e habilidades a serem desenvolvidas [Cotta et al. 2015].

O desafio está em como transformar os alunos em sujeitos ativos, responsáveis pela construção do seu conhecimento. É necessária então a utilização de metodologias que potencializam o pensamento crítico, que desperte a criatividade e a prática de reflexão.

O conhecimento aprendido deve ser lembrado de forma constante para que seja mantido na memória de trabalho, para que possa ser utilizado em atividades de interpretação e de raciocínio lógico [Rosa e Galvão 2015]. Cotta et al. (2015) também enfatiza que os alunos devem saber usar a informação e o conhecimento adquirido na aplicação de situações novas e complexas, de forma autônoma, sabendo “Fazer”, “Conviver” e “Ser”, como mostra a Figura 1.

¹ <https://goo.gl/7fbGe7>

O domínio das competências está ligada, como mostra a Figura 1, em habilidades cognitivas tal como memória, atenção e percepção, mas também em habilidades não cognitivas, como autonomia, estabilidade, perseverança e habilidade em trabalhar em grupo. O aluno, para ter a habilidade de tomada de decisão frente a situações inovadoras, deve compreender como agregar essas competências, conhecimentos e habilidades e assim assumir o papel de criador de soluções relevantes para as situações que lhe são apresentadas.

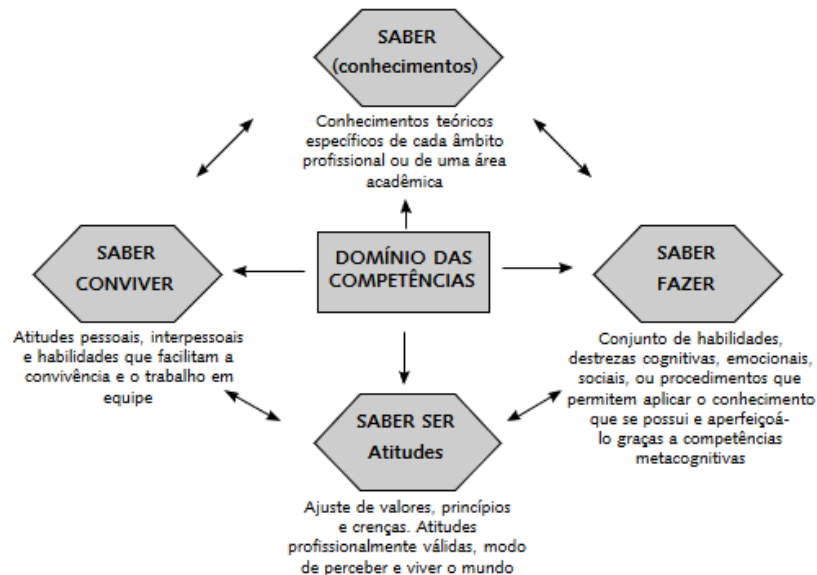


Figura 1. Domínio das competências a um aprendizado constante. Fonte: [Cotta et al. 2015]

1.2.1. Habilidades cognitivas e não cognitivas

O desenvolvimento de habilidades cognitivas e sociais tem como base os processos de aprendizagem, os quais se evidenciam por mudanças relativamente permanentes nos conhecimentos ou comportamentos e ações das pessoas, mudanças essas devidas à experiência, ou seja, às relações sociais e objetivas que os indivíduos experimentam em sua história de vida.

De acordo com Gatti (2005), as habilidades cognitivas são capacidades que fazem o indivíduo competente e que lhe permitem interagir simbolicamente com seu meio ambiente. Essas habilidades formam a estrutura fundamental do que se poderia chamar de competência cognitiva da pessoa humana permitindo discriminar entre objetos, fatos ou estímulos, identificar e classificar conceitos, levantar/construir problemas, aplicar regras e resolver problemas. Elas estão na base dos processos de transferência que propiciam a construção continuada da estruturação de processos mentais cada vez mais complexos na direção da construção/reconstrução de estratégias cognitivas.

Não basta dominar Português e Matemática, por exemplo, se o indivíduo não souber se relacionar com os outros, não for determinado e não conseguir controlar suas emoções, entre outras características da personalidade. Competências socioemocionais, também chamadas competências não cognitivas, de caráter ou qualidades pessoais são o tipo de habilidade envolvida na obtenção de objetivos, no trabalho em grupo e no

controle emocional. São exemplos de competências não cognitivas, como chamam os educadores: autonomia, estabilidade emocional, sociabilidade, capacidade de superar fracassos, curiosidade, perseverança.

O Instituto Ayrton Senna e a Organização para Cooperação e Desenvolvimento Econômico (OCDE) fizeram um estudo com uma ferramenta desenvolvida para a medição de competências socioemocionais [OCDE 2015]. Esse estudo agrupou as habilidades cognitivas e não cognitivas mais importantes para serem avaliadas, como mostra a Figura 2.

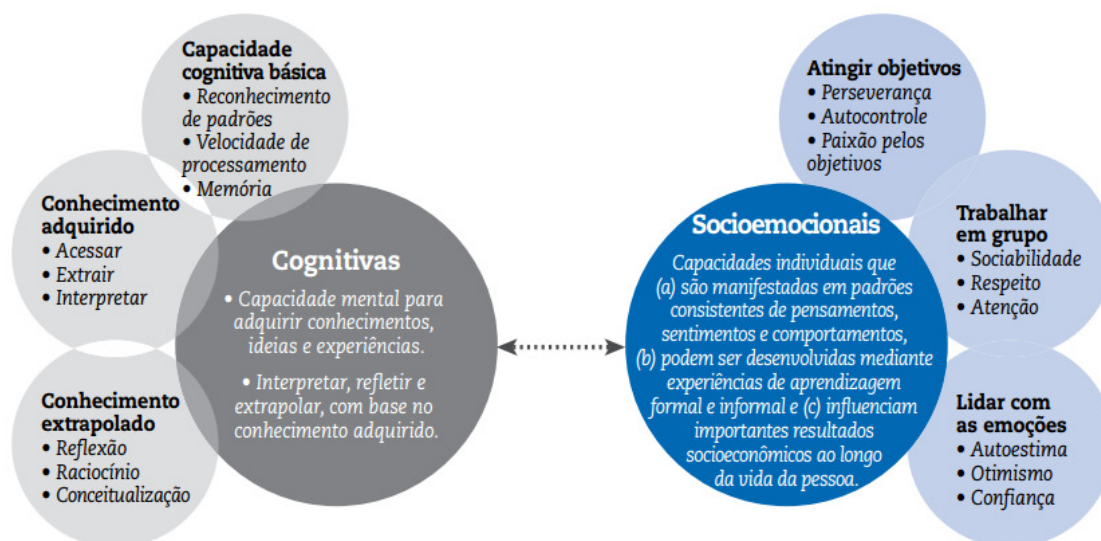


Figura 2. Descrição das habilidades cognitivas e socioemocionais. Fonte: [OCDE 2015].

A avaliação de habilidades cognitivas, geralmente utilizando-se testes padronizados em áreas específicas, foi tida como a mais importante na predição educacional, até devido por sua maior facilidade de medição. Porém, a avaliação das habilidades socioemocionais continua um desafio, pois os alunos ainda estão em fase de desenvolvimento de tais habilidades. O que deve ser ressaltado é que a avaliação deve considerar os dois tipos de habilidades para medir o desempenho do aluno [Silva Junior 2017].

1.2.2. Habilidades Curriculares

Na educação formal é exigida do professor a verificação e mensuração do aprendizado do aluno, apresentando quantitativamente os resultados da aprendizagem. E esses, por sua vez, são obtidos aplicando provas e testes regulares nas instituições do governo e programas e ações do Ministério da Educação brasileiro².

O Indicador de Desenvolvimento da Educação Básica (IDEB) é calculado a partir dos dados sobre aprovação escolar, obtidos no Censo Escolar, e das médias de desempenho nas avaliações do Inep (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira), o Sistema de Avaliação do Ensino Básico (SAEB), composto por: a Avaliação Nacional da Educação Básica (Aneb), a Avaliação Nacional do Rendimento Escolar (Anresc), conhecida como Prova Brasil, e a Avaliação Nacional

² Ministério da Educação, avaliações da aprendizagem: <https://goo.gl/JtVGcU>

da Alfabetização (ANA) e o Exame Nacional do Ensino Médio (ENEM).

Os programas de avaliação implementados pelo governo brasileiro têm chamado a atenção para uma dimensão humana com uma longa tradição de pesquisa na Psicologia: as habilidades cognitivas. Comissões compostas para elaboração dessas provas, os especialistas definem um conjunto de habilidades consideradas essenciais e, a partir delas, constroem os instrumentos de avaliação. As avaliações estruturam-se a partir de matrizes de competências e habilidades que fundamentam a construção dos itens e informam como os autores entendem a inteligência e quais, entre as suas dimensões, devem ser privilegiadas pelo Ensino Básico e Médio.

Um exemplo dessas matrizes são os Parâmetros Curriculares Nacionais (PCN), que definem o conjunto progressivo de habilidades essenciais que todos os alunos devem desenvolver ao longo das etapas e modalidades da Educação Básica, indicando conhecimentos e competências que se espera que todos os estudantes desenvolvam ao longo da escolaridade.

1.2.3. Habilidades para o século XXI

Um comitê formado por educadores, psicólogos e economistas pesquisou sobre o que se espera que os estudantes alcancem nos seus ciclos escolares, nos seus futuros trabalhos e em outros aspectos da vida [Council et al. 2013]. O resultado foi que as “habilidades do século XXI” são habilidades advindas dos desafios impostos ao estilo das novas tecnologias de informação e comunicação (TIC). As tecnologias estão cada vez mais presentes em sala de aula; o professor tem que se preocupar em preparar o aluno para ser atuante em um mundo em transformação e ajudá-lo a desenvolver as competências específicas para o século 21, definindo o que pode oferecer para que crianças e jovens tenham acesso à melhor formação possível [Perrenoud and Thurler 2009].

Segundo Council et al. (2013), as competências foram divididas em três grandes domínios. O primeiro deles é o cognitivo, que é aquele que envolve estratégias e processos de aprendizado, criatividade, memória, pensamento crítico; é o que está relacionado à aprendizagem mais tradicional. O intrapessoal tem relação com a capacidade de lidar com emoções e moldar comportamentos para atingir objetivos. Já o interpessoal envolve a habilidade de expressar ideias, interpretar e responder aos estímulos de outras pessoas.

As competências cognitivas estão agrupadas em três: I - Processos cognitivos e estratégias, tendo como principais termos: pensamento crítico, resolução de problema, análise, raciocínio / argumentação, interpretação, tomada de decisão, aprendizagem adaptativa; II - Conhecimento, tendo como principais termos: alfabetização em informação (pesquisa usando evidências e reconhecendo fontes confiáveis), alfabetização em tecnologia da informação e das comunicações, comunicação oral e escrita; escuta ativa; III - Criatividade e inovação.

As competências intrapessoais estão divididas em mais três grupos: I - Aperfeiçoamento intelectual, tendo como principais termos: iniciativa, autodirecionamento, responsabilidade, perseverança, produtividade, autorregulação tipo 1 (habilidades metacognitivas, incluindo previsão, desempenho e autorreflexão), profissionalismo, ética, integridade, orientação profissional; II - Auto avaliação do Núcleo Positivo: Autorregulação tipo 2 (automonitoramento, autoavaliação,

autorreforço), saúde física e psicológica.

Por fim, as competências interpessoais estão particionadas em dois grupos: I - Habilidades sociais como trabalho em equipe e colaboração: comunicação, colaboração, trabalho em equipe, cooperação, coordenação, empatia, confiança, orientação de serviço, resolução de conflitos, negociação; e II - Habilidades sociais persuasivas como: Liderança, responsabilidade, comunicação assertiva, autorrepresentação, influência social com os outros.

1.3. Jogos Digitais

O trabalho pedagógico a partir dos jogos contribui para que o exercício e o desenvolvimento dos aspectos cognitivos se tornem mais lúdico e prazeroso, ao mesmo tempo em que usufrui das reconhecidas contribuições que o jogo oferece ao desenvolvimento infantil [Kessler et al. 2010]. Os jogos digitais podem ser definidos como ambientes simulados atrativos e interativos que capturam a atenção do jogador ao oferecer desafios que exigem níveis crescentes de destreza e habilidades.

Os jogos promovem o desenvolvimento intelectual, já que para vencer os desafios o jogador precisa elaborar estratégias e entender como os diferentes elementos do jogo se relacionam, desenvolvendo várias habilidades cognitivas, como a resolução de problemas, tomada de decisão, reconhecimento de padrões, processamento de informações, criatividade e pensamento crítico [Gros 2003]. Savi e Ulbricht (2008) elencaram benefícios que os jogos digitais educacionais podem trazer aos processos de ensino e aprendizagem e desenvolvimento de habilidades cognitivas e não cognitivas:

- Efeito motivador: conseguem provocar o interesse e motivam estudantes com desafios, curiosidade, interação e fantasia.
- Facilitador do aprendizado: auxiliam, por exemplo, o entendimento de ciências e matemática quando se torna difícil manipular e visualizar determinados conceitos, como moléculas, células e gráficos matemáticos.
- Desenvolvimento de habilidades cognitivas: promovem o desenvolvimento intelectual, pois em diversos momentos o jogador precisa elaborar estratégias e entender como os diferentes elementos do jogo se relacionam para vencer os desafios.
- Aprendizado por descoberta: com *feedback* instantâneo e o ambiente livre de riscos, os jogadores desenvolvem a capacidade de explorar, experimentar e colaborar.
- Experiência de novas identidades: oferecem aos estudantes oportunidades de novas experiências de imersão em outros mundos e a vivenciar diferentes identidades.
- Socialização: os jogadores também podem servir como agentes de socialização à medida que aproximam os alunos jogadores, competitivamente ou cooperativamente, dentro do mundo virtual ou no próprio ambiente físico de uma escola ou universidade.

Tavares et al. (2013), cita jogos usados no processo de ensino-aprendizagem na disciplina de química como o Jogo Adivinhas da Tabela Periódica, composto por charadas sobre os elementos da tabela periódica. As charadas são feitas e o jogador clica no elemento químico que se encontra logo abaixo, como mostra a Figura 3. Para responder às charadas o jogador possui um tempo, sendo que para cada resposta errada

dada, o jogador será penalizado com o acréscimo de 20 segundos do tempo total. O aluno pode ter auxílio da própria tabela periódica que se localiza na parte inferior à direita da tela. Nesse jogo existe também módulo de palavras cruzadas. O aluno clicar em um dos números para acessar pistas para a palavra correspondente. Escreve-a no retângulo (em letras maiúsculas e com acentos) e clica em “Responder” para que a resposta surja na palavra cruzada. Se o jogador tiver dúvidas pode clicar no botão “Ajuda”, para serem adicionadas sucessivamente as letras da palavra no retângulo. Os autores indicam que a realização de uma avaliação pode ser feita por turma em que o grupo vencedor é aquele que terminar em menos tempo.

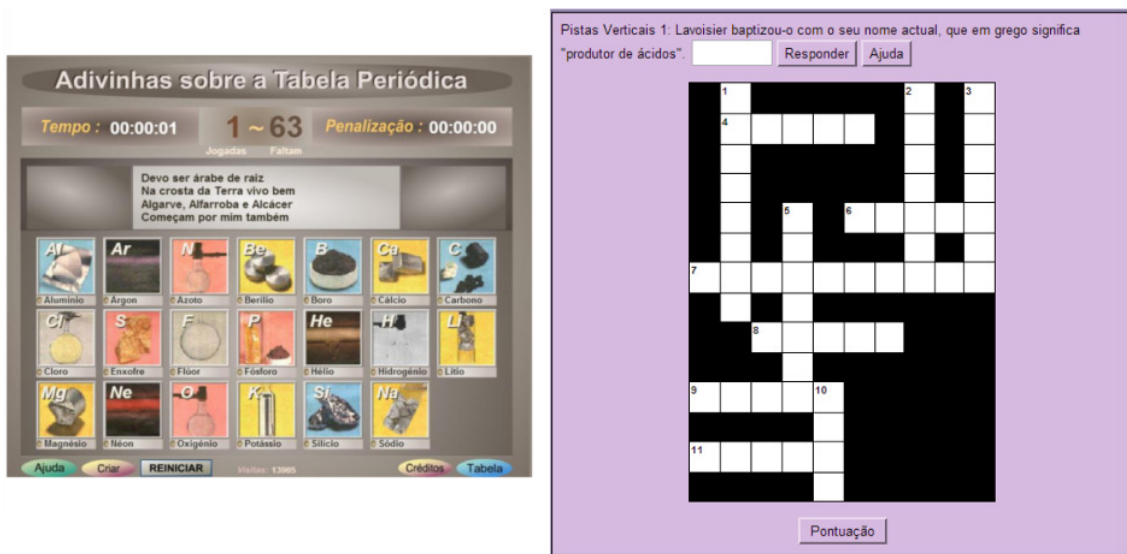


Figura 3. Jogo Adivinhas e Palavras cruzadas sobre a Tabela Periódica. Fonte: [Tavares et al. 2013].

Crayon Physics Deluxe [Kao et al. 2017] é um jogo digital em que os participantes resolvem enigmas físicos com desenhos 2D. Este jogo não só permite aos jogadores resolver os enigmas, mas também aplicar conceitos de física para criar soluções para os enigmas de jogo, permitindo aos alunos aprender conceitos de física a partir da perspectiva do aprendizado baseado em problemas, bem como demonstrar a possível criatividade no design de soluções para o jogo. O trabalho buscou mensurar a aquisição de conhecimento de conceitos físicos (na forma de mapas conceituais) e criatividade de design (na forma de projetos de fases de jogos).

Para passar cada nível de jogo, o jogador move uma bola vermelha, através do vale/cânion para a estrela, desenhando objetos e usando estratégias envolvendo conceitos físicos (Figura 4). Os desenhos são automaticamente transformados em objetos físicos no jogo para resolver o enigma. À medida que um aluno se move de um nível para outro, a experiência em resolver tais enigmas é acumulada e pode ser usada em níveis subsequentes. Com isso, é demonstrada a flexibilidade na criação das soluções dos alunos e aplicação de conceitos físicos para resolver os enigmas.

O jogo Virtual Age [Cheng et al. 2015] auxilia os alunos na compreensão da evolução biológica dos animais, medindo correlações entre aprendizagem de conceito de alunos, desempenho nos jogos e comportamentos no jogo. Os conceitos científicos de evolução são realizados, concretizados e gamificados na tentativa de aproveitar o poder dos jogos para aumentar a compreensão estudantil da evolução biológica. No jogo

existem áreas de recursos (que são constituídas por plantas representativas de cada época) onde os jogadores devem usar seus personagens herbívoros para ocupar e produzir biomassa (Figura 5).

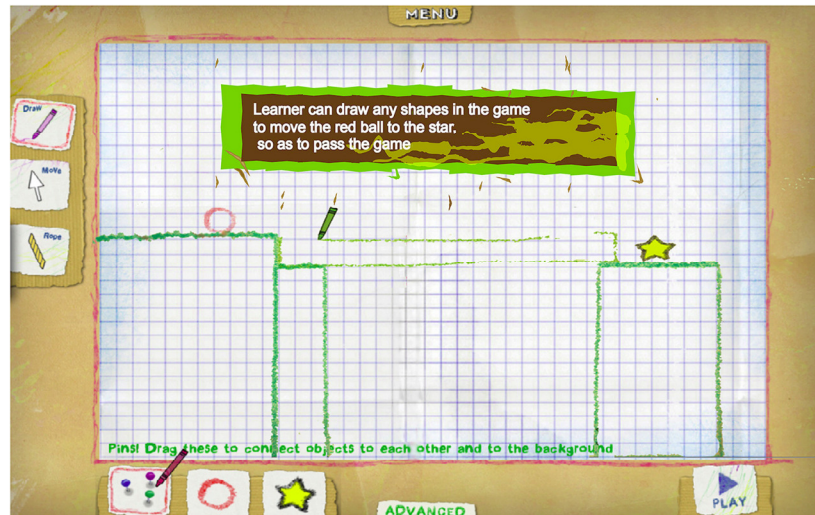


Figura 4. Tela de uma fase do Crayon Physics. Fonte: [Kao et al. 2017]



Figura 5. Uma captura de tela Virtual Age. PC: personagem do jogador; NPC: personagens não jogáveis. Fonte: [Cheng et al. 2015]

O jogador possui o mecanismo de seleção natural, manipulando uma equipe de personagens jogáveis (PCs) para competir com personagens não jogadores (NPCs) controlados pelo computador. Além disso, cada nível (era) tem objetivos para os jogadores completarem. Um sistema de pontuação e um quadro de classificação que avaliam as performances de jogos dos jogadores estão incluídos para aumentar competitividade do jogo. O desempenho de um jogador é representado como uma pontuação mostrada quando o jogo acaba e seu cálculo é feito de forma simples de

acordo com o número e o tipo de caracteres utilizados, se os personagens estão vivos ou mortos, o grau de simbiose alcançado e quantas áreas de recursos estão ocupadas ou não ocupadas por NPCs. Além disso, os pontos de bônus podem ser obtidos se ataques contínuos no mesmo NPC forem realizados.

Physics Playground (PP) de Shute et al. 2015 é um jogo educacional 2D que foi desenvolvido para medir a aprendizagem dos conceitos de física (leis da força e movimento de Newton Massa, gravidade, potencial e energia cinética e conservação do impulso) para alunos da 8ª e 9ª série. As fases do PP exigem que o aluno guie uma bola verde para um balão vermelho. A principal maneira de mover a bola é criando objetos simples ou “agentes de força e movimento”, como eles são chamados no jogo (ou seja, rampas, alavancas, pêndulos e trampolins), desenhados com linhas coloridas usando o mouse, que “Ganham vida” na tela.

Uma rampa é qualquer linha desenhada que ajuda a guiar a bola em movimento (por exemplo, para colocar a bola sobre um buraco). Uma alavanca gira em torno de um ponto fixo e é útil quando um aluno quer mover a bola verticalmente. Um pêndulo de balanço direciona um impulso tangente à direção do movimento, e normalmente é usado para exercer força horizontal, como mostra a Figura 6. Um trampolim armazena energia potencial elástica fornecida por um peso decrescente e é útil quando um aluno quer mover a bola verticalmente.



Figura 6. Solução usando um pêndulo em um problema no Physics Playgrounds. Fonte: [Shute et al. 2016]

O trabalho [Shute et al. 2016] utilizou um jogo chamado Use Your Brainz (versão modificada de Plants vs. Zombies 2), para medir as habilidades de resolução de problemas dos estudantes do ensino médio. No jogo, os jogadores devem plantar uma variedade de plantas especiais em seu gramado para evitar que os zumbis entrem em sua casa (Figura 7). Cada uma das plantas têm atributos diferentes. Por exemplo, algumas plantas (ofensivas) atacam zumbis diretamente, enquanto outras plantas (defensivas) retardam os zumbis para dar ao jogador mais tempo para atacar os zumbis. Algumas plantas geram “sol”, um recurso no jogo necessário para produzir mais plantas. O desafio do jogo consiste em determinar quais plantas usar e onde colocá-las para vencer todos os zumbis em cada nível do jogo.



Figura 7. Tela do Jogo Use Your Brainz. Fonte: [Shute et al. 2016]







	<p>Scene 1: Emotion Regulation Degree to which the child can control what he does and says in an emotionally charged situation</p> <p>Task: Bullies in hall are trying to force the child to pay a toll before allowing him to go to his class. Child can choose to be passive (P), aggressive (AG), or assertive (AS) in interactions with the bullies.</p>
	<p>Scene 2: Impulse Control Degree to which the child can stay focused and on-task during problem solving</p> <p>Task: Teacher instructs the child to feed the elephant before the class can go to recess. The child must gather information from the clipboard (which provides feeding instructions) and then from the teacher in order to accomplish this task.</p>
	<p>Scene 3: Communication Degree to which the child is able to communicate clearly and respectfully, as well as listen accurately in conversation with another character</p> <p>Task: Note on the door indicates the child's class will return at 2:30. Bell rings and the hall monitor informs her that she must get to class. The child must communicate with the hall monitor to find out where her class is and get a hall pass to go and find them.</p>
	<p>Scene 4: Empathy Degree to which the child demonstrates caring for and understanding of another character</p> <p>Task: The child is playing 4-square on the playground and there's a boy on sidelines clearly upset. The child needs to understand what is upsetting this boy in a caring and thoughtful way.</p>
	<p>Scene 5: Cooperation Degree to which the child demonstrates team work with another student in order to solve a problem</p> <p>Task: The child needs to figure out how to catch the parrot flying around the classroom. In order to be successful, the child must work cooperatively with another character.</p>
	<p>Scene 6: Social Initiation Degree to which the child initiates social play appropriately with other characters</p> <p>Task: The child is on the playground and can select to interact with the giraffe or join a 4-square game with three other characters.</p>

Figura 8. Fases e objetivos do Zoo U. Fonte: [Shute et al. 2016]

Zoo U [DeRosier and Thomas 2017] é um mundo escolar virtual que tem por objetivo apresentar situações sociais virtuais que as crianças experimentam na escola primária. O jogo aproveita a propensão natural das crianças em relação aos jogos para

envolvê-los na resolução de problemas sociais e no pensamento crítico para a avaliação baseada no desempenho das habilidades. O presente estudo promove avaliação de habilidades socioemocionais (SE), examinando o grau da competência social para diferenciando os alunos que experimentam problemas sociais de ajuste escolar.

Foram elencadas seis habilidades socioemocionais (SE) que repetidamente demonstraram impacto nos resultados sociais, emocionais, comportamentais e acadêmicos. Essas seis habilidades SE são comunicação, cooperação, regulação emocional, empatia, controle de impulsos e iniciação social. Como mostrado na Figura 8, o Zoo U apresenta crianças com uma série de seis simulações virtuais especificamente criadas para provocar a habilidade social alvo e exigem a aplicação dessa habilidade social para resolver o desafio social apresentado. Coletivamente, essas seis habilidades SE formulam a conceitualização geral da competência social do Zoo U.

O Alien Rescue [Kang et al. 2017] utiliza jogos sérios e de mundo aberto juntamente com a pedagogia de aprendizagem baseada em problemas, em que estudantes com diferentes conhecimentos (por exemplo, especialistas e novatos) usam abordagens para resolver problemas. Os alunos possuem o papel de jovens cientistas que são convidados a se juntarem a uma operação de resgate das Nações Unidas para salvar um grupo de seis alienígenas de uma galáxia distante porque seus planetas foram destruídos. Os alunos estão envolvidos em investigações científicas com um objetivo de encontrar uma casa adequada em nosso sistema solar para mudar cada espécie alienígena.

Uma vez que cada alienígena possui necessidades e características exclusivas, os alunos são desafiados a reunir informações incorporadas em diferentes ferramentas e a integrar essas informações para resolver um problema complexo e mal estruturado para cada alienígena. Os alunos precisam justificar uma solução, fornecendo motivos e evidências. Os alunos precisam explorar os múltiplos espaços funcionais para apoiar processos cognitivos e testes de hipóteses, e desenvolver estratégias para utilizar diferentes ferramentas no jogo. Por meio deste jogo sério, os alunos experimentam processos cognitivos semelhantes à pesquisa científica do mundo real e praticam habilidades cognitivas como geração de hipóteses e resolução de problemas.

Esses exemplos ilustram a abrangência e potencial do uso dos jogos digitais como material de ensino e aprendizagem exercitando diversas habilidades e competências.

1.4. Avaliação de habilidades a partir de jogos digitais

De acordo com Shute e Ventura (2013), existem várias razões para usar jogos na escola: primeiro, como nossas escolas permaneceram praticamente inalteradas por muitas décadas, enquanto o nosso mundo está mudando rapidamente, estamos vendo um número crescente de alunos desengajados; segundo, há uma necessidade urgente de medidas dinâmicas e contínuas de processos de aprendizagem e resultados; terceiro, muitos desses jogos normalmente exigem que um jogador aplique várias competências (por exemplo, criatividade, resolução de problemas, persistência e colaboração) para ter sucesso no jogo.

O trabalho de Serrano-Laguna (2017) realizou um apanhado de técnicas e padrões para sistematizar a análise de aprendizagem em jogos. Em geral, os jogos usam

duas estratégias de interação: I - baseada em eventos, onde o jogo registra eventos pré-especificados quando ocorrem; e II - baseado no estado, onde o jogo repetidamente envia o estado do jogo a uma frequência específica. A maioria dos eventos e estados inclui pelo menos dois atributos: um registro de data/hora, representando o momento em que o evento é gerado e um id de usuário, identificando o jogador que origina o evento. Esses dados básicos permitem aos pesquisadores obterem métricas, como número de jogadores, número de vezes que os jogadores acessaram o jogo ou o tempo jogado.

Para registrar conclusão, os jogos usam valores binários (sim ou não) ou o nível de conclusão alcançado (porcentagem) pelos jogadores. Este tipo de interação baseia-se na noção de que um jogador ao completar totalmente ou partes do jogo também está progredindo em direção a um objetivo de aprendizagem. Serrano-Laguna (2017) afirma que as métricas comuns extraídas desses eventos são a quantidade de níveis/fases/cenas completados e o tempo necessário para completar cada um.

Os jogos podem implementar questões de escolhas realizadas pelos jogadores em um determinado contexto. Essas escolhas podem envolver perguntas com múltiplas respostas, sendo elas apresentadas diretamente ao jogador durante a partida ou integradas em uma caixa de diálogo com um personagem não jogável (NPC). Outros jogos trazem escolhas gerais em que o jogador deve selecionar uma ação entre várias opções. Serrano-Laguna (2017) destaca que em muitos casos, quando um jogador faz uma escolha em um jogo sério, ele precisaria aplicar conhecimento específico para tomar a decisão correta. Esse recurso torna a escolha interessante para futuras avaliações. As métricas comuns obtidas a partir dessas interações são: o tempo gasto para tomar decisões e a razão de correção/erro quando as seleções podem ser pontuadas positivamente ou negativamente (perguntas com uma resposta correta).

Os jogos geralmente dependem de variáveis mensuráveis para calcular os resultados de aprendizagem dos jogadores. As variáveis mais comuns são pontuações, número de mortes no jogo e mortes ou moedas coletadas. Todas essas variáveis estão ligadas ao desempenho do jogador e podem revelar o nível de sucesso nos objetivos de aprendizagem envolvidos. Por exemplo, uma pontuação final alta em um jogo sério pode indicar um bom desempenho em relação aos objetivos de aprendizagem do jogo. Serrano-Laguna (2017) relata que as métricas mais comuns extraídas dessa interação são o valor final de pontuação final do jogo e os índices que podem ser obtidos combinando vários deles coletados de diferentes tentativas de jogo.

Existem abordagens onde há a necessidade rastrear eventos específicos desses jogos. Por exemplo, coletar logs de bate-papo de um jogo *multiplayer*, rastrear o número de vezes que um jogador pediu ajuda, coletar informações biométricas usando vários dispositivos externos. Isso nos leva a concluir que, embora possamos identificar eventos e interações comuns, o uso de interações específicas do jogo também é necessário para avaliar o desempenho do usuário em jogos.

No estudo Kao et al. (2017) aplicaram aprendizagem baseada em problemas da teoria da aprendizagem do construtivismo. Os alunos tiveram duas oportunidades diferentes para demonstrar sua criatividade: uma enquanto eles estavam jogando o Crayon Physics Deluxe, e a outra enquanto eles estavam projetando suas próprias fases de jogo, externalizando seus conceitos de física com desenhos.

O *Consensual Assessment Technique* (CAT) [Baer and McKool 2009] foi

utilizada para avaliar a criatividade dos participantes na concepção de suas próprias fases do jogo. O CAT é frequentemente usado para avaliar produtos de criatividade, como obras de arte, poemas e composições. Os dados das pontuações e das fases elaboradas pelos alunos passaram por técnicas estatísticas para controlar variáveis estranhas executando análises de covariância (ANCOVA) testando se determinados conjuntos de dados possuem médias iguais ou não, levando-se em conta a variação dos números em torno da média. Destacaram cinco variáveis dependentes correlacionadas (sensibilidade, fluência, flexibilidade, originalidade e elaboração) para a criatividade do projeto das fases do jogo, aplicaram uma análise de variância multivariada (MANOVA) para produzir um resultado estatístico neste conjunto de variáveis em vez de realizar vários testes individuais.

De uma escola secundária no norte de Taiwan, 126 participantes em quatro classes foram selecionados para participar do experimento. Os alunos que utilizam o jogo seriam significativamente melhores na resolução de pós-teste do que aqueles no grupo que não jogaram. Demonstrou-se que os próprios professores podem adaptar um jogo digital, como o Crayon Physics Deluxe, para cumprir os objetivos de ensino e aprendizado.

Para investigar como os alunos interagem e usam personagens no jogo Virtual Age [Cheng et al. 2015], criou-se uma base de dados MySQL. Os comportamentos do aluno no jogo, incluindo os personagens, o número de vezes que os personagens estão sendo utilizados, a duração e o número de vezes que visualizam as informações de status, a duração da visualização da informação sobre simbioses e de decompositores, O total de rodadas (turnos) sendo reproduzido, a quantidade total de personagens adquiridos e o desempenho do jogo do aluno (pontuação do jogo) são gravados e usados para análise de dados.

Utilizaram análise de *cluster* para investigar padrões de uso de personagens, resultados de aprendizagem de conceitos e a melhor pontuação do jogo. Três clusters principais que produzem resultados mais eficientes foram, portanto, obtidos usando o algoritmo K-means. A ANOVA e a análise de variância multivariada MANOVA foram calculadas para testar ainda mais as diferenças nos resultados de aprendizado conceitual e a melhor pontuação do jogo nos três clusters.

O estudo que usou o Physics Playground (PP) [Shute et al. 2015] focou na modelagem das relações entre estados afetivos, desempenho no jogo, persistência e aprendizado dos princípios da física, no contexto do jogo. A pesquisa usa de medidas comportamentais e observacionais, coletadas enquanto os alunos interagiram com o jogo no laboratório de informática da escola, em vez de medidas de auto-relato e questionários.

Aplicou-se o *Performance-Based Measure of Persistence* (PMP) a fim de medir o esforço que as pessoas exercem em tarefas difíceis, verificando assim, a persistências dos alunos. Os estados afetivos dos alunos e os comportamentos em tarefa foram observados durante suas interações com Physics Playground, usando o Protocolo de Monitoramento Baker-Rodrigo-Ocuppaugh (BROMP), Observadores treinados realizam anotações de comportamento, observando os alunos usando uma técnica de round-robin. Para verificar o aprendizado no jogo, foram realizados pré-teste e pós-teste por formulário e examinando os dados utilizando o ANOVA no pré-teste e dados pós-teste, que indicou um aumento significativo do pré-teste para pós-teste.

No trabalho de Shute et al. 2016, dividiu-se a capacidade de resolução de problemas em quatro grupos primários: (a) análise de dados e restrições, (b) planejamento de uma via de solução, (c) uso de ferramentas e recursos de forma eficaz e eficiente, e (d) monitoramento e avaliação de progresso. O primeiro grupo mapeia a aquisição de “regras” e os restantes para “aplicação de regras”. Depois de definir os grupos da resolução de problemas, identificou-se variáveis relevantes no jogo para cada uma das quatro grupos.

Criou-se uma rede Bayes exclusiva para cada nível de jogo (42 níveis) pois a maioria dos indicadores não se aplicam em todos os níveis e as redes simples tornam os cálculos mais eficientes. As estimativas relacionadas com a habilidade de resolução de problemas de um jogador são atualizadas à medida que as evidências em curso se acumulam a partir de suas interações com o jogo. As redes Bayes acumulam dados das variáveis e propagam esses dados em toda a rede atualizando as distribuições de probabilidade. Desta forma, as variáveis influenciam dinamicamente as estimativas da habilidade de resolução de problemas de um aluno e em cada um dos quatro grupos de resolução de problemas. Por exemplo, um grupo da resolução de problemas é a capacidade de um jogador utilizar as ferramentas de forma eficaz e eficiente. Uma das dezenas de plantas do jogo é chamada *Iceberg Lettuce*, que é usado para congelar temporariamente um zumbi, atrasando assim o ataque do zumbi (Figura 7).

Para validar as estimativas no jogo da habilidade de resolução de problemas, os resultados foram correlacionados com dois testes externos: *Raven's Progressive Matrices* (RPM), um teste que examina a capacidade dos sujeitos de fazerem inferências com base em informações fornecidas, e o MicroDYN que é um sistema de simulação que testa a habilidade dos sujeitos de adquirir e aplicar informações em complexos ambientes de resolução de problemas. Um experimento foi realizado com 55 alunos da 7ª série de Illinois. Testou-se as correlações entre os resultados das Redes Bayesianas de resolução de problemas e as pontuações com os dois testes externos. Os resultados mostraram que a avaliação baseada em jogo de habilidades de resolução de problemas correlacionou-se significativamente com a Raven e MicroDYN. Assim, a avaliação de discriminação por resolução de problemas mostrou-se válida.

No jogo Zoo U [DeRosier and Thomas 2017] desenvolveu-se fases para avaliação utilizando teorias educacionais usando uma abordagem implementada por designers de jogos educacionais conhecidos como Evidence Centered Design (ECD) [Almond et al. 2015]. Definiu-se quais as competências queriam avaliar, bem como os comportamentos reais que proporcionaram evidências para essas competências. Em seguida, desenvolveu-se tarefas no jogo que estavam associadas tanto às competências de interesse quanto aos comportamentos reais relacionados.

À medida que as crianças navegam nas fases do Zoo U, o software rastreia e guarda as seleções de menu acessados e comportamentos no jogo. O algoritmo de pontuação para cada cena é adaptado ao objetivo social dessa cena e as opções de escolha refletem a resolução de problemas sociais se adaptam para a situação social apresentada. Tal como acontece com as situações sociais do mundo real, as escolhas realizadas pelo jogador no Zoo U influenciam como os objetos da cena respondem e quais opções de ações estão disponíveis para a criança. Assim, as pontuações para escolha de menu de ações refletem a qualidade das escolhas individuais do menu, bem como a sequência dessas escolhas.

Exemplos de possíveis índices de comportamento no jogo incluem o número de vezes que um objeto ou personagem particular é clicado, o tempo gasto em uma determinada atividade, a ordem das ações tomadas durante a resolução de problemas e o número de pistas ou dicas dada pelo sistema necessário para resolver o problema. O trabalho realizou um experimento com uma amostra total de alunos composta por 270 alunos, aproximadamente divididos uniformemente entre os gêneros (57% do sexo masculino) com distribuição racial de 58% de brancos, 24% de afro-americanos e 18% de grupos multi-raciais ou de grupos minoritários raciais. Os resultados do estudo indicam que jogos podem ser efetivamente usados para avaliação de habilidades. O uso de simulações virtuais baseadas em jogos para avaliação baseada em desempenho das habilidades sociais emocionais oferece uma série de vantagens em relação aos métodos de avaliação de habilidades mais tradicionais

Em [Kang et al. 2017], são definidos dados de jogabilidade usados para identificar os padrões de comportamento dos alunos à medida que interagem com as ferramentas disponíveis no jogo no Alien Rescue. O jogo registra cada ação à medida que cada aluno interage com o meio ambiente do jogo. Os dados contêm um identificador de estudante, uma ferramenta acessada por um aluno, um tipo de ação que um aluno está tomando (por exemplo, Zoom, Click, Abrir ou Fechar), notas adicionais sobre as interações do aluno e um registro de data/hora para cada ação. O algoritmo cSPADE foi usado para obter padrões sequenciais frequentes.

Realizou-se uma pesquisa com 202 alunos do sexto ano de uma escola secundária na região sudoeste dos Estados Unidos. Este estudo examinou os frequentes padrões sequenciais entre estudantes de baixa e alta performance e mostrou que as estratégias de resolução de problemas eram diferentes entre esses dois grupos de atuação. De acordo com Kang et al. (2017), ao usar este método analítico integrado, podemos obter uma melhor compreensão do caminho de aprendizado das estratégias de desempenho e resolução de problemas dos estudantes com diferentes características de aprendizagem em um contexto de jogo sério.

1.4.1. Uso de ECD e Redes Bayesianas para a modelagem de competências

Stealth Assessment é utilizado no ambiente para apoiar a aprendizagem e avaliação de conteúdos e competências. Representa um processo “silencioso” pelo qual os dados de desempenho do aluno são continuamente recolhidos durante a realização de atividades e inferências são feitas sobre as competências relevantes. Os resultados das inferências são salvas em estados de competência e armazenadas em um modelo dinâmico do aluno. A *Stealth Assessment* destina-se a apoiar a aprendizagem e manter um fluxo, onde o grau de envolvimento dos alunos seja, não por esperar recompensas externas, mas simplesmente para a alegria de fazer. O objetivo é eventualmente desfocar a distinção entre avaliação e aprendizagem, pretendendo remover (ou reduzir seriamente) a ansiedade de teste, sem sacrificar a validade e a confiabilidade [Shute e Ventura 2013].

Dado o objetivo de usar jogos para apoiar a aprendizagem em ambientes escolares e em outros cenários, os avaliadores precisam garantir que as avaliações sejam válidas e confiáveis. Os jogadores devem estar cientes de que a interação deles com os jogos afetam as competências, deixando transparente os resultado de suas avaliações. Uma maneira de atender a esses requisitos é usar *Stealth Assessment* [Shute 2011].

Durante as partidas, os alunos produzem naturalmente ricas sequências de ações ao executar tarefas complexas, baseando-se nas habilidades ou competências que queremos avaliar (por exemplo, habilidades de pesquisa científica e criatividade). As evidências necessárias para avaliar as habilidades são assim fornecidas pelas interações dos jogadores com o próprio jogo, o que pode ser contrastado a uma atividade.

Responder corretamente à pergunta é evidência de que se pode conhecer um certo fato. Mas, ao analisar uma sequência de ações dentro de uma missão em ambiente de jogo, onde cada resposta ou ação fornece evidências incrementais sobre o domínio atual de um fato, conceito ou habilidade específicos, a *Stealth Assessment* pode inferir o que os alunos sabem e não sabem. Shute e Ventura (2013) afirma que métodos para analisar a sequência de comportamentos para inferir essas habilidades não são tão óbvios e sugerem que sejam usadas avaliações baseadas em evidências, por exemplo, o *Evidence Centered Design* [Almond et al. 2015].

As representações de conhecimento são importantes considerando avaliações educacionais porque para várias pessoas e estágios dentro do processo, as representações diferentes da informação serão: algoritmos automatizados para calcular pontuação; representações matriciais de modelos psicométricos para estatísticos; e mapas de domínio para área de conteúdo Especialistas. O framework ECD fornece modelos para as representações de conhecimento, como modelos psicométricos e de tarefas.

Evidence Centered Design

Todas as avaliações educacionais têm em comum o desejo de encontrar o motivo para os alunos dizerem ou fazerem determinadas ações, a inferências sobre o que sabem ou podem fazer, busca encontrar solucionar essa resposta. Métodos de avaliação evoluíram para abordar este problema de uma forma sistemática. Os modelos de medição da teoria clássica de testes e, mais recentemente, a teoria da resposta a itens (IRT), a análise de classes latentes e a modelagem de diagnóstico cognitivo têm se mostrado bastante satisfatórios para testes em grande escala e questionários em sala de aula.

Entretanto, as avaliações os testes padronizados são cada vez mais insatisfatórios para orientar a aprendizagem e avaliar o progresso avaliação dos alunos. Os avanços na tecnologia permitem elaborar provas de conhecimento que capturarem performances mais complexas das respostas. Um dos gargalos mais sérios que enfrentamos, no entanto, é dar sentido a dados complexos desses resultados.

O *Evidence Centered Design* (ECD) estabelece um *framework* conceitual para os elementos de uma avaliação. O *framework* é genérico suficiente para suportar uma ampla variedade de tipos de avaliação, desde os testes e questionários em sala de aula até sistemas para treinamentos e avaliações baseadas em simulação. Segundo Almond et al. (2015), a estrutura de projeto baseia-se nos princípios do raciocínio probabilístico e nas exigências da produção e entrega da avaliação. A concepção da avaliação em tal estrutura garante que a forma como as evidências são recolhidas e interpretadas depende do conhecimento implícito dos alunos e objetivos abordados pela avaliação [Kim et al. 2016], através do esboço das tarefas, captura dos dados durante a realização dessas tarefas e análises dos dados gerados para identificar adequadamente o desenvolvimento de competências de diferentes alunos. A arquitetura definida ajuda na divisão e coordenação entre o trabalho de diferentes especialistas, tais como especialistas em

assuntos, estatísticos, instrutores, autores de tarefas, desenvolvedores e designers de interfaces.

O processo de ECD aborda uma série de perguntas que devem ser feitas em qualquer projeto de avaliação; o que, onde, como estamos medindo e quanto precisamos medir. As respostas a essas perguntas são registradas em vários modelos chamados de competência, evidência, tarefa e *assembly*; a coleção de todos os modelos utilizados para uma determinada avaliação é chamada de *Conceptual Assessment Framework* (CAF) [Kim et al. 2016]. Esses modelos orientam, então, o desenvolvimento do mecanismo operacional de avaliação, tais como tarefas e modelos estatísticos (Figura 9).

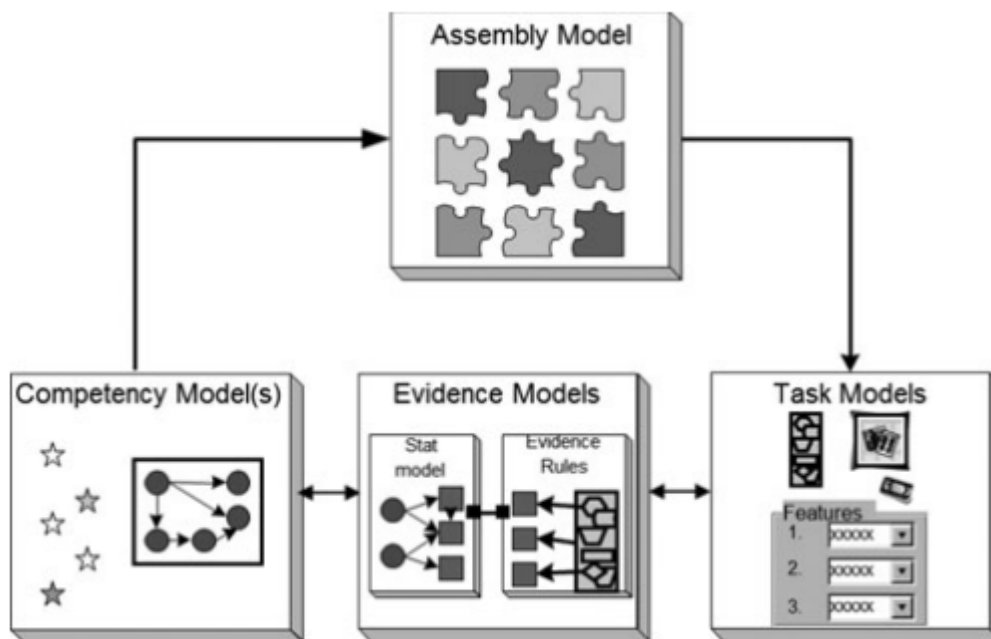


Figura 9. Arquitetura do ambiente de avaliação. Fonte: [Kim et al. 2016]

Os Modelos de Competência (CM) são as informações sobre o que estamos medindo, são as habilidades/competências (estrelas na Figura 9) que o avaliador deseja que sejam adquiridas ou aprimoradas nos alunos ao final da avaliação. Essas competências são variáveis no CM, que descrevem aspectos de habilidades, conhecimentos, traços e atitudes de interesse. A distribuição das variáveis, suas dependências no modelo de competência são descritas usando um modelo estatístico. O ECD permite CMs com relações multidimensionais e as redes bayesianas fornecem uma linguagem gráfica para expressar essas relações [Almond et al. 2015].

Os Modelos de Evidência (EM) são responsáveis por registrarem como medimos as competências; um aluno interagindo com as tarefas produz *work products*. No GBA, o *work products* geralmente consiste em uma transcrição das interações do jogador, geralmente salvas em logs do sistema. O EM fornece regras para atualizar o CM do aluno. Um modelo de evidência possui duas partes: as regras de evidência (o lado direito do EM na Figura 9), onde são definidos os observáveis, características do *work products* que fornecem evidência sobre uma ou mais competências; e o modelo estatístico descreve as relações entre as variáveis de competência e os observáveis, expressando a probabilidade de que um aluno com determinado perfil produza um valor

específico para cada variável observável. No GBA, as regras de evidência são frequentemente expressas em código de computador (por exemplo, o jogador completou com sucesso um nível de jogo? O jogador usou uma ferramenta específica na solução de um problema?).

O Modelo de Tarefa (TM) é encarregado de identificar onde medimos as competências. Uma tarefa é uma unidade de atividade que é tentada pelo aluno, que por sua vez, produz um *work product*. Essa tarefa pode ser tão simples quanto uma pergunta de múltipla escolha, mas o ECD encoraja os designers de avaliação a pensarem em tarefas de desempenho complexas também. No GBA, uma tarefa é uma unidade autocontida do jogo. Em alguns jogos, o jogo é naturalmente dividido em níveis, e as tarefas correspondem aos níveis do jogo. Em outros casos, uma tarefa pode ser um segmento de uma interação mais complexa, e definir o limite da tarefa pode ser desafiador [Kim et al. 2016]. Um modelo de tarefa (TM) descreve uma família de tarefas relacionadas (muitas vezes todas elas são variações em um único protótipo) e as variáveis do modelo de tarefa expressam características de uma tarefa que podem ser manipuladas para criar variantes e mudar a força ou foco de evidência [Almond et al. 2015].

O Modelo *assembly* é responsável por definir o quanto precisamos medir. Um modelo *assembly* (AM) especifica como os CMs, EMs e TMs trabalham juntos para gerar evidências suficientes para formar uma avaliação válida [Almond et al. 2015]. O AM especifica a quantidade de evidências são necessárias para cada variável de competência. O AM garante que os resultados de alunos que pouco interagem com as tarefas se sobressaiam sobre os que interagem mais.

O CM, EMs e AM juntos compõem o modelo de medição da avaliação. Para construir um mecanismo de pontuação, os projetistas devem construir uma relação matemática do CM e um EM para cada tarefa que pode ser selecionada pelo AM. Kim et al. (2016) analisaram a realização desses modelos usando redes bayesianas. O CAF completo contém mais dois modelos adicionais: modelo de sistemas de apresentação que descreve como as tarefas são renderizadas em diferentes plataformas (por exemplo, computador de mesa versus telefone inteligente). E o modelo de entrega descreve outros fatores relacionados ao ambiente de avaliação, incluindo a elegibilidade para participar da avaliação.

Redes Bayesianas

Redes Bayesianas (RB) são modelos probabilísticos aplicados em problemas que envolvem dados incertos e imprecisos, permitindo representar a dependência entre variáveis e são uma forma compacta de representar a distribuição de probabilidade conjunta [Russell et al. 1995]. Uma RB é representada por intermédio de um grafo dirigido acíclico, onde seus nós são variáveis que possuem diferentes estados e probabilidades e as arestas dos nós indicam a relação de dependências, de modo que, o nó de origem da aresta influencia no nó de destino.

Caso um nó Y receba uma aresta de um nó X, ele tem a sua probabilidade definida pela probabilidade condicional $P(Y|X)$. Da mesma forma, se nenhuma variável estende uma aresta para X, este possui apenas as suas probabilidades independentes, ou

a priori, que são definidas apenas por $P(X)$ [Pearl 1988].

Na Figura 10, tem-se um exemplo de um grafo que representa uma rede bayesiana. Os nós círculos representam variáveis latentes, ou seja, variáveis que terão seus valores dados em condição aos seus nós destinos. Os nós em forma de caixa representam as variáveis observadas, valores que podem ser atualizados. As setas (arestas) representam as dependências condicionais. As variáveis dos nós na cauda de uma seta são “pais”, e as variáveis na cabeça da seta são as “filhos”. Por exemplo, B é o pai de X_1 e é filho de A. Assim, um nó pai (origem) tem a sua probabilidade dependente do resultado do seu nó filho (destino),

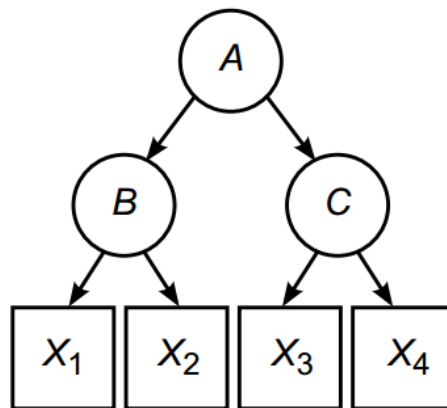


Figura 10. Exemplo de uma rede bayesiana. Fonte: [Culbertson. 2016]

A vantagem de um RB vem das independências condicionais implicadas pelo gráfico e a correspondente simplificação à regra geral de multiplicação. Cada variável é condicionalmente independente de todas as outras variáveis, tendo em conta as variáveis que o cercam [Almond et al. 2007]. A regra geral de multiplicação fornece uma distribuição de probabilidade conjunta como produto de distribuições sucessivas de probabilidade condicional, por exemplo,

$$P(A, B, C, X_1, X_2, X_3, X_4) = P(A) * P(B | A) * P(C | A, B) * P(X_1 | A, B, C) * P(X_2 | A, B, C, D, X_1) * P(X_3 | A, B, C, D, X_1, X_2) * P(X_4 | A, B, C, D, X_1, X_2, X_3)$$

Embora esta representação recursiva possa ser escrita para qualquer ordem de variáveis, a ordem que corresponde às distribuições de probabilidade condicional associadas aos nós do RB permite uma simplificação considerável devido às independências condicionais implícitas no gráfico:

$$P(A, B, C, X_1, X_2, X_3, X_4) = P(A) * P(B | A) * P(C | A) * P(X_1 | B) * P(X_2 | B) * P(X_3 | C) * P(X_4 | C)$$

Além disso, as distribuições podem ser especificadas simplesmente tomando recursivamente o produto da distribuição condicional de um nó e as distribuições condicionais dos pais, por exemplo:

$$P(X_3) = P(X_3 | C) * P(C | A) * P(A)$$

Assim, as relações complexas entre muitas variáveis podem ser descritas por meio de relações condicionais entre subconjuntos de variáveis muito menores, que podem ser lidas diretamente do gráfico [Culbertson 2016].

Para realizar as inferências, as probabilidades são atualizadas a partir das

variáveis observáveis, denominadas evidências, com estados possuindo probabilidade de 0% a 100%. Cada nova evidência gera um impacto nas probabilidades da rede [Millan et al. 2010]. Elas podem ser utilizadas de duas formas: predição que consiste em aplicar as evidências na rede e observar os comportamentos das variáveis que são influenciadas por elas e diagnóstico aplicando a evidência na variável influenciada e observando qual a variável exerceu maior influência sobre ela.

Em Seffrin et al. (2013) tem-se um exemplo do funcionamento de uma rede bayesiana. A rede representa a influência de rinite e resfriado sobre a ocorrência de espirros de uma pessoa. A existência de rinite é influenciada devido a pessoa ser alérgica e ocorra a presença de gatos no local, indicando a possibilidade de alergia a gatos. Os arranhões nos móveis na casa são influenciados diretamente pela existência de gatos na casa. As tabelas apresentam as chances de cada evento ocorrer dadas as probabilidades para cada nó da rede (Figura 11).

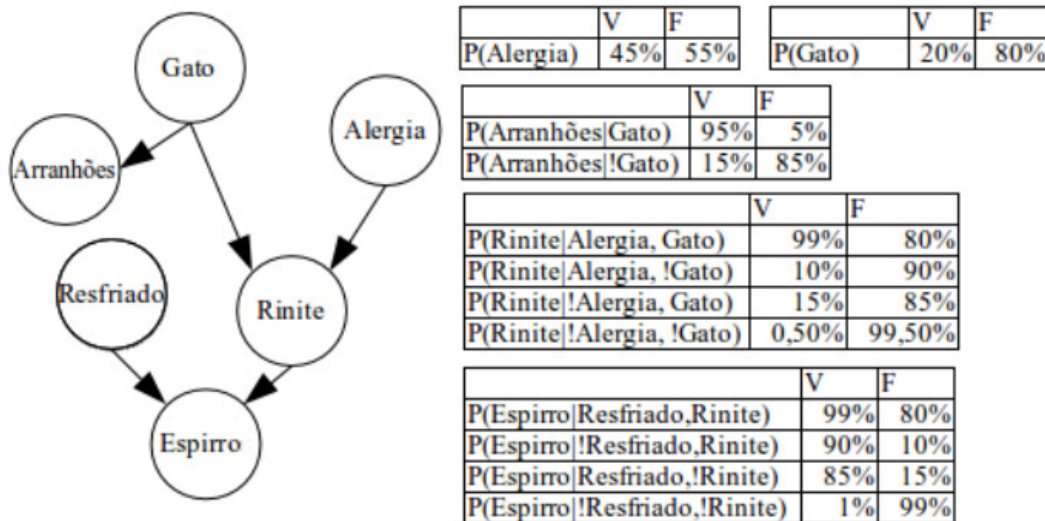


Figura 11. Exemplo de uma rede bayesiana. Fonte: [Seffrin et al 2013].

Em aplicações de RB para a avaliação educacional, Culbertson (2016) afirma que as variáveis latentes representam características cognitivas do domínio de conteúdo e as folhas geralmente representam variáveis observadas (embora possam ser conectadas entre si para codificar dependências não relacionadas às latentes). As variáveis observadas podem incluir itens de teste tradicionais (por exemplo, escolha múltipla) ou outras medidas com base no desempenho do examinado, como a saída de rotinas de processamento acústico em uma tarefa oral de decodificação de palavras [Tepperman, Lee e Alwan, 2011], interações com um ambiente de pesquisa científica baseado em computador [Ting e Phon-Amnuaisuk 2012], ou ações em jogos de computador relevantes para o ensino [Shute 2005]. Os itens podem estar conectados a um ou mais nós latentes, semelhante à Q-matrix no DCM [Almond, 2010; Tatsuoka, 1983].

Junior et al. (2014) elaboraram uma Rede Bayesiana para análise do desempenho da aprendizagem de estudantes considerando recursos pedagógicos do AVA. A estrutura da rede proposta é construída de forma automatizada e permite uma análise que contempla dois enfoques: avaliação geral do desempenho do estudante e outro, específico, no desempenho de cada questão considerada no AVA. O trabalho

proposto por [Seffrin and Jaques 2015] consiste na implementação de uma RB para o mapeamento do conhecimento de álgebra elementar, mais especificamente, de operações relacionadas a resolução de equações de 1º grau. O artigo [Silveira et al. 2012] apresenta uma abordagem para a representação hierárquica do conhecimento para a avaliação do aluno em lógica proposicional. A avaliação hierárquica consiste em avaliar o estado do estudante de conhecimento em vários níveis de granularidade.

Modelagem de competências por meio de Redes Bayesianas

Construir uma Rede Bayesiana para modelagem de competências envolve duas etapas principais. Primeiro, as variáveis psicométricas são definidas por especialistas do conteúdo, definem as competências específica e as variáveis observáveis e, em seguida, desenham um grafo dirigido acíclico que captura as relações de independência condicional importantes. Em segundo lugar, para cada variável no modelo, a equipe de projeto especifica a probabilidade condicional dessa variável dada seus pais (ou seja, nós a partir dos quais uma flecha se estende conectando-os ao nó filho) no grafo.

Shute et al. (2016) elaboraram um modelo de competência de resolução de problemas, a partir de uma revisão da literatura chegaram as quatro facetas (habilidades) que o aluno deve desenvolver: (a) análise de dados e restrições, (b) planejamento de uma via de solução, (c) uso de ferramentas e recursos de forma eficaz e eficiente, e (d) monitoramento e avaliação de progresso. Depois de definir indicadores para as facetas da resolução de problemas, os autores identificaram as variáveis observáveis no jogo para cada uma delas.

Um RB revela graficamente as dependências condicionais que existem entre as várias variáveis na rede. Consiste em variáveis do modelo de competência (isto é, resolução de problemas e suas quatro facetas) e os observáveis associados (indicadores) que estão estatisticamente ligados às facetas. À medida que as evidências em curso se acumulam a partir de suas interações com o jogo, as estimativas relacionadas com a habilidade de resolução de problemas de um jogador são atualizadas.

Para encontrar essas variáveis observáveis, foram identificadas ações observáveis (indicadores) no jogo que proporcionam evidências para as facetas. Para isso, Shute et al. (2016) realizaram um processo iterativo que começou por fazer um brainstorming de uma lista de potenciais ações derivadas de jogar e uma série de soluções para alguns níveis.

Após de listar todos os potenciais indicadores, os autores avaliaram suas (a) relevância para sua faceta associada, e (b) viabilidade de ser implementado no jogo. Removeram indicadores que não estavam intimamente relacionados às facetas ou eram muito difíceis ou vagos de implementar no jogo. Repetiram o processo de adição, avaliação e exclusão de indicadores até encontrar uma lista satisfatória, chegando a 32 indicadores: 7 para analisar dados e restrições, 7 para planejar uma via de solução, 14 para usar ferramentas e recursos efetivamente e 4 para monitorar e avaliar o progresso. Exemplos de indicadores para cada faceta são mostrados na Tabela 1.

Uma vez que o conjunto de variáveis observáveis foi determinado (incluindo indicadores positivos e negativos), o próximo passo é estabelecer relações estatísticas aceitáveis entre cada nível, variáveis observáveis as variáveis do modelo de

competência. As regras de pontuação basearam-se na coleta de instâncias relevantes das ações observáveis no jogo e, em seguida, em uma classificação em categorias discretas, como sim / não (um aluno fez ou não fez alguma ação no nível), ou pobre / ok / bom / muito bom (Dependendo da qualidade das ações). Construiu-se Redes Bayesianas para acumular os dados recebidos e atualizar as variáveis observáveis em relação aos níveis de competência.

Facetas	Indicadores de Facetas
Análise de dados e restrições	Plantar > 3 <i>Sunflowers</i> antes da chegada da segunda onda de zumbis
	Seleciona as plantas antes da barra de zumbis encher.
Planejamento de uma via de solução	Coloca produtores de sol atrás, plantas ofensivas no meio e plantas defensivas na frente
	Plantas Twin <i>Sunflowers</i> ou <i>Sunflowers</i> em níveis que exigem a produção de X sol
Uso de ferramentas e recursos de forma eficaz e eficiente	Usar plantas alimentos quando há > 5 zumbis no quintal ou os zumbis estão chegando perto da casa (dentro de 2 quadrados)
	Danos > 3 zumbis ao disparar um <i>Coconut Cannon</i>
Monitoramento e avaliação de progresso	Arrancar <i>Sunflowers</i> e as substituir por plantas ofensivas quando a proporção de zumbis para plantas exceder 2:1

Tabela 1. Exemplos de indicadores para cada faceta de resolução de problemas. Fonte: [Shute et al. (2016)].

Por exemplo, uma faceta da resolução de problemas é a capacidade de um jogador utilizar as ferramentas de forma eficaz e eficiente. Caso o aluno realize uma ação listada como mau uso da ferramenta, o arquivo de log captura as informações de posicionamento e se comunica com o modelo de evidência sobre o uso ineficaz da ferramenta, que por sua vez atualiza as estimativas sobre o estado atual da habilidade para resolução de problemas do aluno.

Para registrar essas informações é criada uma lista com as informações das facetas, o indicador e o evento do jogo que gera informações sobre esses indicadores, como pode ser visto na Tabela 2.

Faceta	Indicador	Indicadores Relevantes	Evento	Especificação do evento
Uso efetivo / eficaz de ferramentas	37	O jogador planta iceberg lettuce dentro do alcance de um ataque de snapdragon (espaço de 3? 3 na frente de um snapdragon) [R]	Indicator_planted_iceberg_in_snapdragon_range	Quando o jogador prepara uma iceberg lettuce, verifique os tiles próximos para um snapdragon. Razão = o número de iceberg lettuce em tiles próximos de um snapdragon / o número de iceberg lettuce plantadas. Ratio para Estado: $0 \leq x \leq 0,25$ "muito bom" $0,26 \leq x \leq 0,50$ "bom" $0,51 \leq x \leq 0,75$ "ok" $0,76 \leq x \leq 1,0$ "pobre"
	12	Usar alimento para as plantas quando houver < 3 zumbis na tela.	Indicator_percent_low_danger_plant_food_usage	Razão = # de alimento da planta usado quando há <3 zumbis na tela / número total de alimentos vegetais usados. Ratio para Estado: $0 \leq x \leq 0,25$ "muito bom" $0,26 \leq x \leq 0,50$ "bom" $0,51 \leq x \leq 0,75$ "ok" $0,76 \leq x \leq 1,0$ "pobre"

Tabela 2. A comunicação entre o log do jogo e os nós de rede Bayes (facetas e indicadores). Fonte: [Shute et al. (2016)].

Uso efetivo / eficaz de ferramentas	Muito Bom	Bom	Ok	Pobre
Alto	0.53	0.32	0.11	0.04
Medio	0.36	0.36	0.21	0.07
Baixo	0.19	0.32	0.31	0.18

Tabela 3. CPT para o indicador #37 no nível 7. Fonte: [Shute et al. (2016)].

As relações estatísticas envolvendo o indicador #37 e sua variável de competência associada (uso efetivo / eficiente da ferramenta) são definidas pela probabilidade em uma tabela de probabilidade condicional (CTP). Por exemplo, a Tabela 3 mostra a tabela de probabilidade condicional para o indicador #37 no nível 7.

Por exemplo, o valor de 0,53 na primeira célula indica que, se o jogador estiver (teoricamente) alto no uso efetivo/eficiente de ferramentas, a probabilidade é de 0,53 que ele ou ela cairá no estado “muito bom” do indicador #37. Após os dados de desempenho (evidência) chegarem no log, as estimativas sobre a capacidade de usar ferramentas serão atualizadas com base no teorema de Bayes.

Os autores definiram as probabilidades iniciais nas RBs com base no pressuposto de que os alunos teriam uma probabilidade igual de serem altos, médios ou baixos em termos de habilidades de resolução de problemas. Isso significa que cada aluno começa com o mesmo modelo de estudante inicial. No entanto, quando as evidências do desempenho do aluno individual entra na rede, as estimativas se tornam progressivamente mais precisas - refletindo o verdadeiro status do aluno sobre a competência.

Em Shute et al. (2016) vemos a Redes Bayesianas criada com as informações das Tabelas 1 e 2, Figura 12, para medir as habilidades de resolução de problemas de um aluno: o nó principal; os quatro nós representando os quatro grupos analisados: análise de dados e restrições, planejamento de solução, uso de ferramentas e recursos de forma eficaz e eficiente, e monitoramento e avaliação de progresso e; dois indicadores relacionados ao uso efetivo de ferramentas (I12 e I37). Todas as evidências recebidas de um aluno em um indicador servem para atualizar as estimativas sobre os nós vinculadas. Por sua vez, a evidência é propagada em toda a rede. Este processo produz uma RB atualizada por aluno para cada nível que eles desempenham.

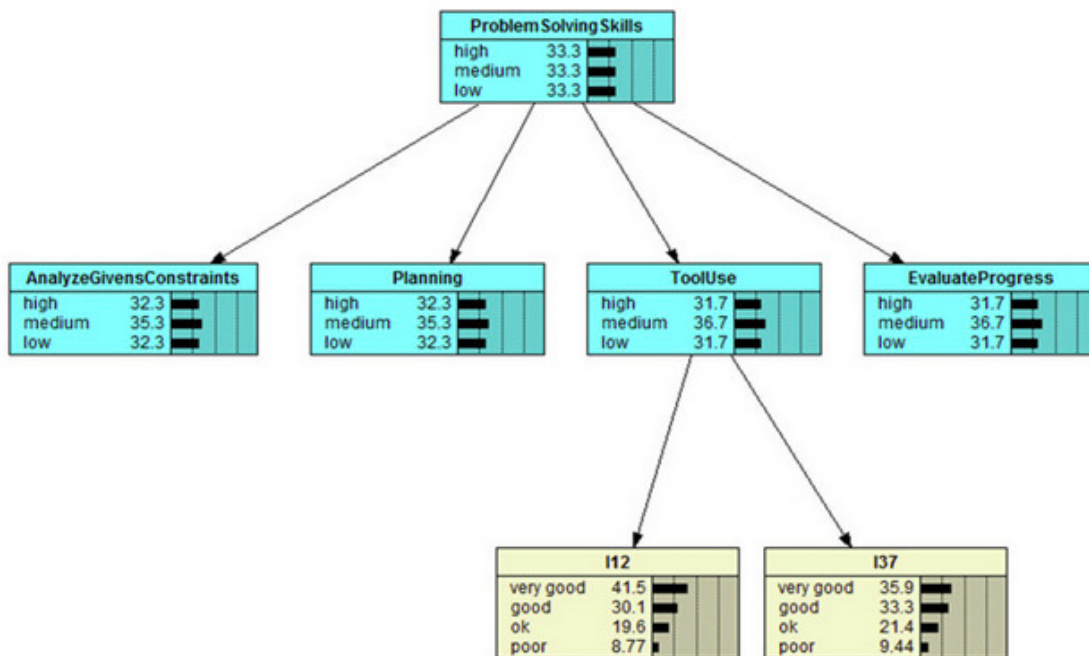


Figura 12. Rede Bayesiana de um aluno gerada ao jogar o Use Your Brainz.
Fonte: [Shute et al. 2016]

De modo geral, os trabalhos citados acima justificam o método pela necessidade de um diagnóstico preciso e flexível de habilidades do aluno em um determinado domínio. Para isso, é modelada a propagação da avaliação de um nível específico de

conteúdo de conhecimento para níveis mais gerais, utilizando inferência Bayesiana.

ECD na prática

O procedimento de 10 passos permite que designers e desenvolvedores de jogos e professores criem, desenvolvam e revisem sistemas, aplicando-os para qualquer avaliação baseada em jogo usando redes Bayes como o mecanismo de pontuação.

A Figura 13 fornece uma visão geral das dez etapas estabelecidas e aplicadas no desenvolvimento do Physics Playground (PP) [Shute et al. 2015]. Estas etapas são organizadas em quatro fases distintas: (a) criação de níveis em PP (Criação de Tarefas), (b) criação de modelos de avaliação (CAF), (c) colocação de todos os modelos de avaliação usando redes Bayes como máquinas de avaliação (por exemplo, Nuts & Bolts), e (d) avaliação e refinamento dos modelos de avaliação (ou seja, Depuração). Embora a Figura 13 indique um fluxo linear através das fases, em sua implementação real o processo deve ser iterativo, conforme indicado pelas setas tracejadas. Dessa forma, problemas que podem aparecer em etapas posteriores exigem muitas vezes a revisão das decisões de projeto feitas anteriormente.

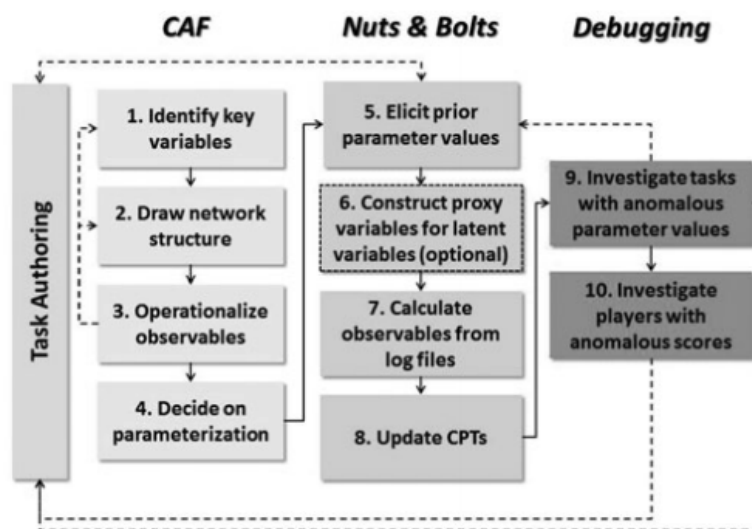


Figura 13. Passos para a construção de redes bayesianas em avaliação baseada em jogos. Fonte: [Shute et al. 2016]

O primeiro passo é identificar as variáveis: (a) determinar os objetivos e propósitos da avaliação, (b) definir as competências de interesse em relação à avaliação, (c) descrever a relação(s) entre cada variável de competência e que podem ser medidos no jogo, e (d) identificar observáveis dentro do jogo que podem fornecer provas para uma ou mais competências.

No PP, usou-se uma Q-matrix [Shute et al. 2015] para documentar a relação entre variáveis observáveis associadas a tarefas e variáveis de competência. Foram adicionadas informações detalhadas sobre tarefas e quais objetos no jogo geram as evidências. A Q-matrix combina informações dos modelos de evidência (quais habilidades são aproveitadas) e o modelo assembly (quantas tarefas há de cada tipo). Outra função da Q-matrix é verificar se tivemos provas suficientes para cada variável de proficiência.

O segundo passo é desenhar a estrutura da Rede Bayesiana contendo as variáveis de competência e de tarefa com as informações levantadas no primeiro passo, com isso temos identificado os modelos competências (CM) e evidências (EM) do ECD. A Figura 14 mostra a rede desenvolvida para o PP, o nó Newton's Three Laws modela a dependência entre o conhecimento dos quatro agentes de movimento (*Ramp Knowledge*, *Lever Knowledge*, *Pendulum Knowledge* e *Springboard Knowledge*) e pode servir como uma pontuação geral para relatórios. O Modelo de Evidências (EM) possui quatro variáveis observáveis, uma para cada agente de movimento. Estes poderiam assumir os valores de Ouro (o jogador criou uma solução elegante/eficiente), Prata (o jogador resolveu o nível, mas não eficientemente) e Nenhum (o jogador não resolveu o nível). Dependendo de qual agente o jogador usou para resolver o problema, a variável observável apropriada seria definida e os outros não seriam observados. A Figura 14 mostra os quatro nós da base do EM e os sete nós superiores mais claros do CM.

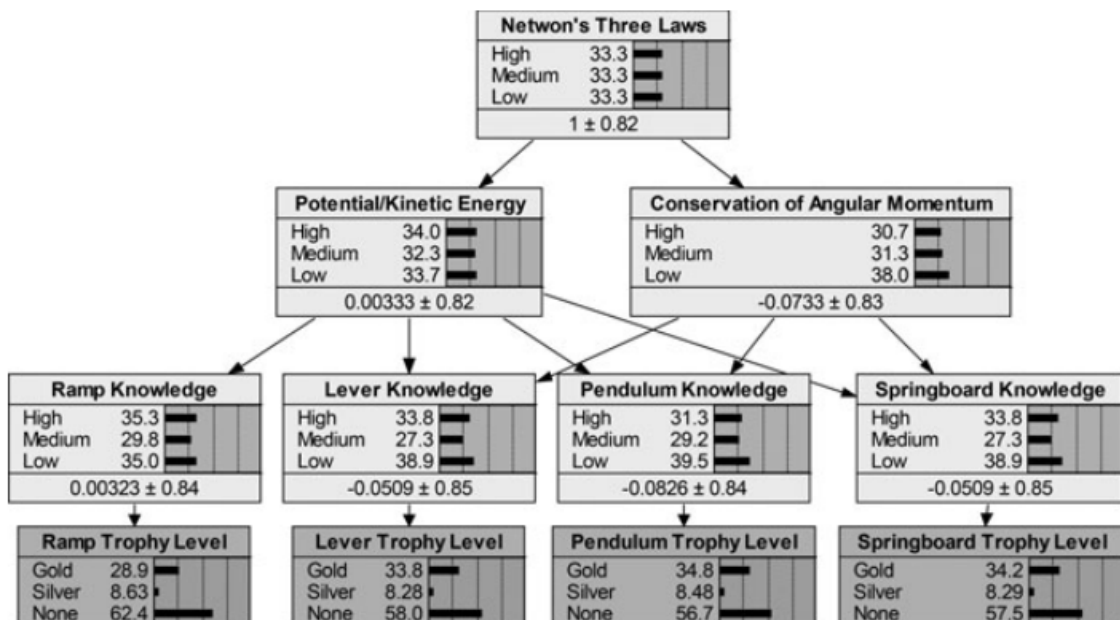


Figura 14. Rede Bayesiana com distribuições condicionais do jogo PP. Fonte: [Shute et al. 2015]

No terceiro passo são criadas as definições operacionais para variáveis observáveis. A principal fonte de evidência é a interação dos jogadores com o jogo que são salvos em arquivos de log. No PP, o arquivo de log contém uma descrição de todos os objetos desenhados pelo jogador, tempo necessário para desenhar objeto, tamanho, forma e localização, bem como informações sobre o nível como um todo (se um troféu de ouro ou prata foi ganho, a duração do tempo gasto no nível e quantas vezes o jogador reiniciou o nível).

No quarto passo é realizada a parametrização de tabelas de probabilidade condicional (CPT). O PP concede um troféu de prata para qualquer solução para um nível no jogo e troféu de ouro por uma solução eficiente. Portanto, os observáveis são variáveis categóricas ordenadas naturalmente (com categorias de ouro, prata e nenhuma).

No quinto passo são configuradas as probabilidades a priori. O PP consiste em 74 níveis diferentes, cada um dos quais possuía quatro variáveis de resultados

observáveis (correspondentes aos quatro agentes de movimento). Para cada variável de resultados, um designer de jogos atribuiu quatro valores: dois valores de dificuldade entre -5 e 5 para os troféus de ouro e prata para esse nível e dois valores de discriminação que variam entre 0 e 2. Esses valores são ajustados levados em consideração cada nível do jogo e cada variável observada possui diferentes valores, onde os valores mais altos indicam que ela deve ser usada, pois está presente na solução mais sofisticada. Por exemplo, uma determinada fase possui uma dificuldade de nível médio e a intenção é que os jogadores usem uma solução de trampolim, que já está presente no nível para impulsionar a bola verde para o Balão.

Dadas estas estimativas de parâmetros, consegue-se calcular os CPTs para as quatro variáveis observáveis. Para isso foi desenvolvido um pacote de software R [Almond, 2015]. Este processo foi repetido para todos os 74 níveis de jogo. Desse modo, produziu um conjunto completo de redes Bayesianas (ou seja, o CM e um EM por nível)

A sexta etapa é opcional, visa construir variáveis de proxy para variáveis latentes. Caso seja utilizada outra fonte de recurso para as variáveis observáveis, como pré-teste quanto o pós-teste, como uma primeira passagem para determinar dificuldades, dividimos as pontuações em valores altos, médios e baixos. O uso das variáveis de proxy no lugar das variáveis latentes permitiu o uso do pacote de software R [Almond, 2015] para realização do passo anterior.

O sétimo passo é calcular as variáveis observáveis dos arquivos de log do jogo. A pontuação de uma avaliação consiste em duas fases: identificação de evidências (EI) - calcular as evidências observáveis do *work product* utilizando dados processados (*raw data*) a partir dos dados (*Log Files*) gerados pelo jogo (*Game Engine*), gerando informações que serão adicionadas na rede; acumulação de evidências (EA) - atualiza as variáveis de competência com base nos dados das variáveis resultados observados no EI. Em PP, o EA é um script que roda em cima das redes Bayesianas. Esse processo, pode ser visto na Figura 15.



Figura 15. identificação e acumulação de evidências em PP. Fonte: [Shute et al. 2015]

Oitavo passo é atualizar as CPTs. Caso os dados de algumas tabelas sejam em muitos níveis, pode haver um ou mais agentes raramente utilizados em soluções. Shute et al. (2015) indicam algoritmos de EM juntamente com o AM, que alternam entre os cálculos fazendo com que os valores não fiquem desbalanceados, ou seja, que o resultado das probabilidades leve em consideração a quantidade de evidências inseridas no sistema.

O nono e décimo passos são dedicados à depuração. Ao completar o quarto

passo, produziu-se um par de dificuldade e parâmetros de discriminação para cada nível de jogo e agente. O primeiro nível de depuração foi procurar valores de dificuldade ou discriminação excessivamente baixos e altos. Para casos com valores de parâmetro incomuns, o próximo passo foi examinar os CPT estimados e as contagens de troféus por habilidade relevante. Em alguns casos, não houve problema. Porém, Shute et al. (2015) afirmaram se o CPT parecer estranho, deve-se revisar o nível do jogo.

Constrói-se planilhas de evidências para cada aluno. As planilhas de evidências são uma representação gráfica de como a probabilidade de um alto nível de “Newton’s Three Laws” muda à medida que a evidência de cada nível de jogo é registrada no sistema. Saltos repentinos na probabilidade podem indicar um possível problema. Níveis com altos pesos de evidências podem revelar problemas com a estratégia de jogo de um jogador.

Esses exemplos ilustram como analisar, criar e configurar as Redes Bayesianas, indicando o uso de eventos no jogo para enviar informações sobre as interações do usuários. Mostrou também passos de como utilizar e projetar os componentes do ECD para que com o uso das Redes Bayesianas possam inferir o conhecimento dos alunos sobre determinadas habilidades.

Os passos do ECD que foram listados devem ser seguidos para que o ambiente seja implementado com sucesso. Em Shute et al. (2016), usaram o do ECD para quantificar a habilidade de resolução de problemas. Ao serem ativados, os eventos que foram identificados nas fases do jogo *Use your Brainz* enviam informações sobre as interações do aluno para o servidor onde implementou-se o ambiente do ECD. Essas evidências recebidas sobre o aluno, são coletadas de forma contínua pelos logs do jogo e servem para atualizar as estimativas sobre suas facetas (s) vinculadas produz uma RB atualizada por aluno para cada nível que eles desempenham.

Depois de desenvolver os BNs (um para cada nível no jogo) e integrá-los ao código do jogo, pode-se adquirir estimativas em tempo real dos níveis de competência de cada jogador no nó principal (habilidade de resolução de problemas) e suas facetas constituintes.

Shute et al. (2016) explica se um jogador que plantou consistentemente uma Iceberg Lettuce na frente de um Snapdragon dentro de um determinado nível no jogo, suponha que a proporção final de Iceberg Alface plantada em frente de Snapdragons ao número total de Iceberg Lettuces plantadas foi de 88%. Esse valor seria categorizado no estado mais baixo do nó (ou seja, "pobre" no indicador #37) e as evidências seriam inseridas e propagadas em toda a BN.

A rede, neste momento, estimaria que o jogador provavelmente apresentaria pouca utilização efetiva de ferramentas: $Pr(\text{uso de ferramentas} = \text{baixa evidência}) = 0,61$ e, portanto, relativamente baixo na habilidade geral de resolução de problemas: $Pr(\text{Resolução de problemas} = \text{baixa evidência}) = 0,50$ (Figura 16).

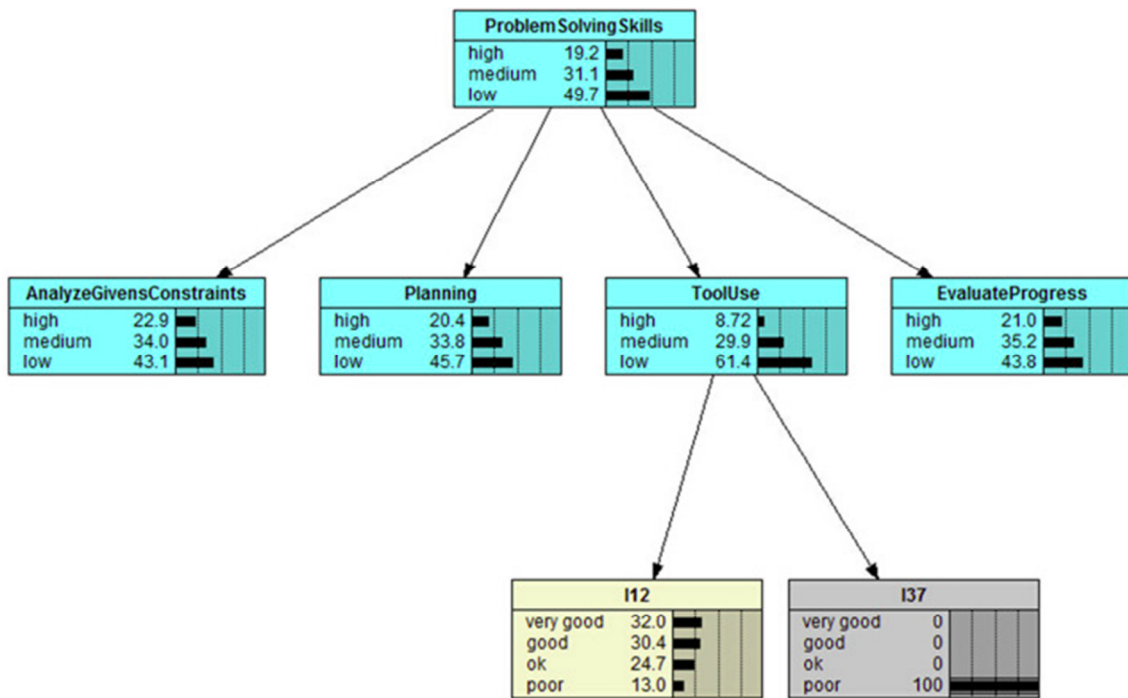


Figura 16. Evidência de mau uso de uma ferramenta recebida pela Rede Bayesiana. Fonte: [Shute et al. 2016]

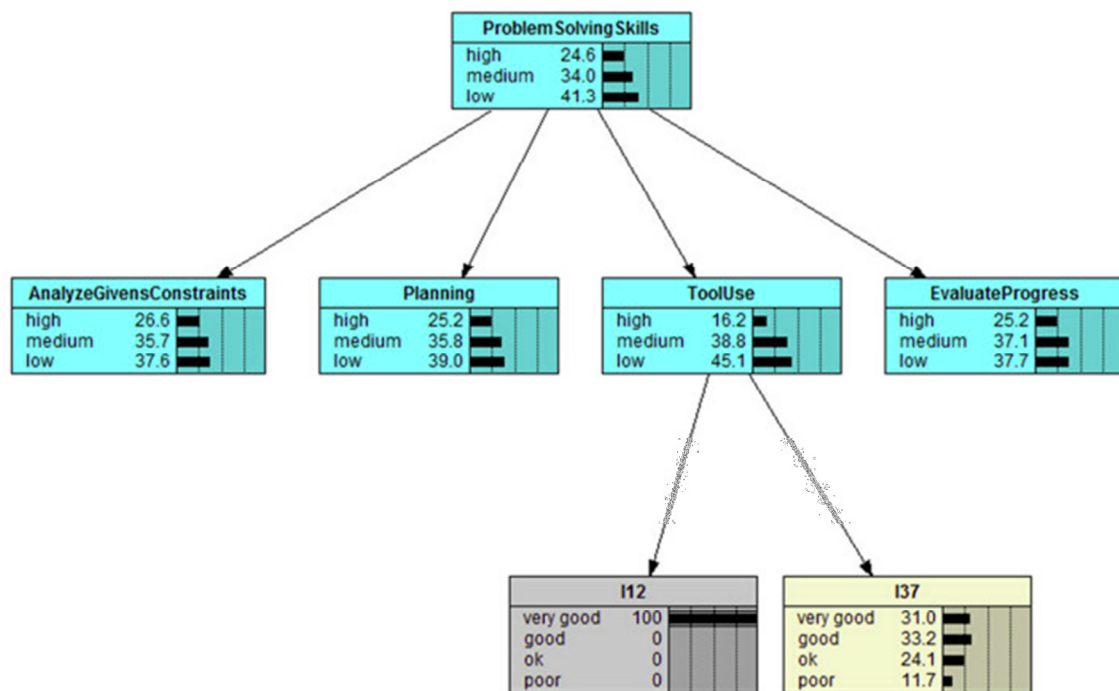


Figura 17. Atualização de RB quando os alimentos são efetivamente utilizados. Fonte: [Shute et al. 2016]

Durante outra partida, o jogador identificou e corrigiu corretamente o erro de plantação de uma Iceberg Lettuce perto de Snapdragons, focando em expandir o poder de Snapdragons, fortalecendo-os com alimentos para aumentar o poder da planta. No entanto, o alimento vegetal é um recurso muito limitado, e se ele é usado

prematuramente, é um desperdício. Usar os alimentos somente quando há vários zumbis presentes sugerem que o jogador entende a função dos alimentos e percebe que o alimento vegetal é um recurso escasso que deve ser conservado para situações críticas (indicador # 12 - veja a última coluna na Tabela 2 para a regra de pontuação associada).

O RB incorpora esta nova evidência e atualiza as estimativas das competências do jogador dos últimos estados mostrados na Figura 17. A distribuição de probabilidade do nível de uso efetivo da ferramenta do jogador é: Pr (uso de ferramentas = baixo) = 0,45, que diminuiu (ou seja, o jogador melhorou) do último estado com base no desempenho anterior relativo ao indicador #37. Pr (uso de ferramentas = médio) = 0,39, que aumentou do último estado (novamente, o jogador está melhorando); Pr (uso de ferramentas = alto) = 0,16, que também aumentou de antes. A distribuição de probabilidade para a habilidade geral de resolução de problemas do jogador mostra o mesmo padrão de melhoria que o uso da ferramenta.

1.5. Considerações Finais

Avaliar a aprendizagem significa determinar as competências adquiridas por meio de aspectos quantitativos e qualitativos. Na educação formal é exigida do professor a verificação e mensuração do aprendizado do aluno, apresentando quantitativamente os resultados da aprendizagem. E esses, por sua vez, são obtidos por intermédio de provas e testes, que na maioria das vezes não contribui para construção do conhecimento do aluno.

É necessário repensar a forma como é realizado o ensino, propondo diferentes formas de avaliação, diversificando os recursos utilizados, levando em consideração a cultura e os meios de expressão que permeiam os indivíduos presentes nesse processo e oferecendo novas alternativas para os indivíduos interagirem e se expressarem. Nesse contexto, a introdução de atividades lúdicas como jogos digitais com finalidades pedagógicas são uma metodologia para ensinar de forma motivadora, possibilitando uma prática significativa daquilo que está sendo aprendido e potencializando a construção e avaliação do conhecimento.

Ao mesmo tempo que o aluno aprende jogando, a metodologia *Stealth Assessment* é utilizada no ambiente para avaliação de conteúdos e competências, capturando os dados de desempenho do aluno durante a realização de atividades e realizando inferências sobre as competências relevantes. Os resultados das inferências são salvos, dessa forma, removendo (ou reduzindo seriamente) a ansiedade dos alunos nos teste.

Realizou-se uma pesquisa de técnicas e padrões para sistematizar a análise de aprendizagem em jogos utilizados na literatura como eventos e estados, escolhas realizadas pelos jogadores, variáveis mensuráveis, bem como essas estratégias estão inseridas em jogos.

É indicado como implementação da *Stealth Assessment* o uso de avaliações baseadas em evidências, como o ECD que se baseia nos princípios do raciocínio probabilístico e nas exigências da produção e entrega da avaliação, através do esboço das tarefas, captura dos dados durante a realização dessas tarefas e análises dos dados gerados para identificar adequadamente o desenvolvimento de competências de diferentes alunos.

Para implementar o ECD em jogos, pesquisadores utilizam passos que auxiliam os designers, desenvolvedores de jogos e professores a criarem, desenvolverem e revisarem sistemas, desde identificação das variáveis de competência e de tarefa, estrutura da Rede Bayesiana, parametrização de tabelas de probabilidade condicional (CPT), configurar as probabilidades a priori, calcular as variáveis observáveis, atualizar as CPTs e por fim depuração para identificar anomalias nos resultados.

Com isso, o trabalho mostrou uma avaliação formativa com o intuito de melhorar o processo de ensino-aprendizagem com o auxílio do uso de informações geradas através da interação dos alunos com os jogos. Por intermédio dessa avaliação, informações sobre o desenvolvimento do aluno são fornecidas ao professor servindo de instrumento de avaliação contínua.

Referências

- Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J. D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement, 44*(4), 341-359.
- Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., and Williamson, D. M. (2015). An introduction to evidence-centered design. pages 19–40. Springer.
- Baer, J., & McKool, S. S. (2009). Assessing creativity using the consensual assessment technique. In *Handbook of research on assessment technologies, methods, and applications in higher education* (pp. 65-77). IGI Global.
- Cheng, M. T., Lin, Y. W., & She, H. C. (2015). Learning through playing Virtual Age: Exploring the interactions among student concept learning, gaming performance, in-game behaviors, and the use of in-game characters. *Computers & Education, 86*, 18-29.
- Culbertson, M. J. (2016). Bayesian networks in educational assessment: The state of the field. *Applied Psychological Measurement, 40*(1), 3-21.
- Cotta, R. M. M., Costa, G. D e Mendonça, E. T. (2015). Portfólios críticos-reflexivos: uma proposta pedagógica centrada nas competências cognitivas e metacognitivas. *Interface (Botucatu), 19*(54):573-88.
- Council, N. R. et al. (2013). Education for life and work: Developing transferable knowledge and skills in the 21st century. National Academies Press.
- DeRosier, M. E., & Thomas, J. M. (2017). Establishing the criterion validity of Zoo U's game-based social emotional skills assessment for school-based outcomes. *Journal of Applied Developmental Psychology*.
- Gatti, B. (2005). Habilidades cognitivas y competencias sociales. *Enunciación, 10*(1):123–132.
- Gros, B. (2003). The impact of digital games in education. *First Monday, 8*(7):6–26.
- Junior, Ademar Crotti, Beatriz Wilges, and Silvia Modesto Nassar. "Modelagem bayesiana da aprendizagem de estudantes em um AVA." *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*.

Vol. 25. No. 1. 2014.

- Kang, J., Liu, M., & Qu, W. (2017). Using gameplay data to examine learning behavior patterns in a serious game. *Computers in Human Behavior*, 72, 757-770.
- Kao, G. Y. M., Chiang, C. H., & Sun, C. T. (2017). Customizing scaffolds for game-based learning in physics: Impacts on knowledge acquisition and game design creativity. *Computers & Education*.
- Kessler, M. C., de Paula, C. G., Albe, M. H., Manzini, N., Barcellos, C., Carlson, R., Marcon, D., and Kehl, C. (2010). Impulsionando a aprendizagem na universidade por meio de jogos educativos digitais. In Simpósio Brasileiro de Informática na Educação - SBIE, volume 1.
- Kim, Y. J., Almond, R. G., and Shute, V. J. (2016). Applying evidence-centered design for the development of game-based assessments in physics playground. *International Journal of Testing*, 16(2):142–163.
- Millan, E., Loboda, T., and Perez-de-la Cruz, J. L. (2010). Bayesian networks for student model engineering. *Computers & Education*, 55(4):1663–1683.
- OCDE, O. p. C. e. D. E. (2015). Skills for social progress: The power of social and emotional skills. Disponível em <<https://goo.gl/YnywhV>>. Acesso em: 22 mai. 2017.
- Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA
- Perrenoud, P. and Thurler, M. G. (2009). *As competências para ensinar no século XXI: a formação dos professores e o desafio da avaliação*. Artmed Editora.
- Prensky, M. (2012). *Aprendizagem baseada em jogos Digitais*. SENAC SP.
- Rosa, G. D. A. e Galvão, A. C. T. (2015). Conhecimento prévio e aprendizagem no ensino: implicações à luz do efeito reverso da *expertise* e de construtos computacionais da cognição. *Ciências & Cognição*; Vol 20(2) 229-237.
- Russell, S., Norvig, P., and Intelligence, A. (1995). *A modern approach*. Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs, 25:27.
- Savi, R. and Ulbricht, V. R. (2008). Jogos digitais educacionais: benefícios e desafios. *Revista Novas Tecnologias na Educação*, 6(2):10.
- Seffrin, H. and Jaques, P. (2015). Avaliando o conhecimento algébrico dos estudantes através de redes bayesianas dinâmicas. In Simpósio Brasileiro de Informática na Educação - SBIE, volume 26, page 987.
- Seffrin, H. M., Rubi, G., & Jaques, P. (2013). Uma Rede Bayesiana aplicada a Modelagem do Conhecimento Algébrico do Aprendiz. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)* (Vol. 24, No. 1, p. 597).
- Sena, S., Schmiegelow, S. S., Prado, G. M. B. C., Sousa, R. P. L. e Fialho, F. A. P. (2016). Aprendizagem baseada em jogos digitais: a contribuição dos jogos epistêmicos na geração de novos conhecimentos. *Revista Novas Tecnologias na Educação*. V. 14, No 1.
- Serrano-Laguna, Á., Martínez-Ortiz, I., Haag, J., Regan, D., Johnson, A., & Fernández-

- Manjón, B. (2017). Applying standards to systematize learning analytics in serious games. *Computer Standards & Interfaces*, 50, 116-123.
- Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction*, 55(2):503–524.
- Shute, V. and Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. MIT Press.
- Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., ... & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers & Education*, 86, 224-235.
- Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, 63, 106-117.
- Silva Junior, W. S. (2017). Evidências sobre habilidades cognitivas e competências socioemocionais dos alunos em idade escolar: formação, desenvolvimento e o papel da Escola no Brasil. Tese de doutorado. Programa Programa de Pós-Graduação em Desenvolvimento Econômico. Universidade Federal do Pará.
- Silveira, S. R., Rangel, A. C. S., and de Lima Cir´ iaco, E. (2012). Utilização de jogos digitais para o desenvolvimento do raciocínio lógico-matemático. *Tear: Revista de Educação, Ciência e Tecnologia*, 1(1).
- Tatsuoka, K. K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354. doi:10.1111/j.17453984.1983.tb00212.x
- Tavares, R., Souza, R. O. O., and Correia, A. d. O. (2013). Um estudo sobre a “tic” e o ensino da química. *GEINTEC-Gestao, Inovacao e Tecnologias* , 3(5):155–167.
- Tepperman, J., Lee, S., & Alwan, A. (2011). A generative student model for scoring word reading skills. *IEEE Transactions on Audio, Speech, and Language Processing*, 19, 348-360. doi:10.1109/TASL.2010.2047812
- Ting, C.-Y., & Phon-Amnuaisuk, S. (2012). Properties of Bayesian student model for INQPRO. *Applied Intelligence*, 36, 391-406. doi:10.1007/s10489-010-0267-7
- Zieky, M. J. (2014). An introduction to the use of evidence-centered design in test development. *Psicología Educativa* 20, 79-8.

Sobre os autores

Handerson Bezerra Medeiros

Possui graduação em Análise e Desenvolvimento de Sistemas pelo IFRN (2011) e mestrado em Engenharia de Software pela UFRN (2013). Atuou como docente substituto da UFRN (2017), atuando principalmente nos seguintes temas: lógica de programação, Desenvolvimento para ambiente desktop, Estrutura de dados, Programação concorrente e distribuída, Programação distribuída, Sistema de informação, Administração de Banco de Dados, Física aplicada a Jogos, Desenvolvimento de Jogos para dispositivos móveis, Inteligência Artificial para Jogos, Desenvolvimento de Jogos Web. Integrante do grupo de pesquisa GameEdu, no Instituto Metr pole Digital (IMD-UFRN) desenvolvendo pesquisas em jogos educacionais com foco em minera o de dados.  reas de atua o: Ci ncia da computa o, com  nfase em an lise e desenvolvimento de sistemas, Sistemas Distribu dos e Jogos Digitais.

Eduardo Henrique da Silva Aranha

Doutor (2009) em Ci ncia da Computa o pelo CIn/UFPE, tem experi ncia como professor no ensino superior desde 2002, bem como experi ncia na ger ncia e desenvolvimento de software desde 1997. Atualmente   professor do Departamento de Inform tica e Matem tica Aplicada da UFRN e coordena o Laborat rio de Pesquisa em Games e Educa o do Instituto Metr pole Digital da UFRN. Atualmente realiza pesquisas principalmente no uso de games para melhoria da educa o b sica e em engenharia de software experimental. Atua em parceria com outros centros de pesquisa, escolas do ensino b sico e empresas de jogos digitais e de educa o adaptativos que auxiliem no fortalecimento de habilidades e compet ncias lingu sticas e matem ticas de crian as e adolescentes do ensino fundamental.

Isabel Dillmann Nunes

Possui gradua o em Ci ncia da Computa o pela UFPEL (1998), mestrado em Ci ncia da Computa o pela UFRGS (2001) e doutorado em Ci ncia da Computa o pela UFCG (2014). Atua nas  reas de Engenharia de Software, mais especificamente com Requisitos de Software e Modelagem de Sistemas e na  rea de Inform tica Educacional, com foco em Learning Analytics e Design Instrucional. Atualmente atua no Instituto Metr pole Digital ministrando disciplinas de Design de Software, Processo de Desenvolvimento de Software Educativo, Introdu o  s T cnicas de Programa o no Curso de Bacharelado em Tecnologia da Informa o e Engenharia de Requisitos no Mestrado Profissional de Engenharia de Software e as disciplinas de Sistemas Colaborativos e An lise de Aprendizagem no Mestrado Profissional em Inova o e Tecnologias Educacionais.