

ELO e TRI: Estimando a Habilidade dos Estudantes em uma Plataforma Online de Programação

Fabiana Zaffalon, FURG, fabinhazaffalon@gmail.com
André Prisco, FURG, prisco.c3@gmail.com
Ricardo de Souza, FURG, rcrdsou@hotmail.com
Rafael dos Santos, FURG, rapennas@gmail.com
Neilor Tonin, UNICER, ntonin@gmail.com Jean
Luca Bez, UNICER, jeanbez@uricer.edu.br Silvia
Botelho, FURG, silviacb.botelho@gmail.com

Resumo: Métodos de avaliação da proficiência dos estudantes vem ganhando destaque nos últimos anos. Existe um número crescente de cursos *online* e plataformas que disponibilizam repositórios de questões ou exercícios onde os métodos de avaliação ocorrem de forma automática. Esse trabalho faz uma análise dos dados gerados através de dois modelos que tem por objetivo estimar a habilidade dos estudantes: ELO e TRI. O ELO foi desenvolvido para classificar jogadores através do histórico de jogo enquanto a TRI estima a habilidade através de um conjunto de respostas dadas a um conjunto de itens. Utilizamos uma base de dados disponibilizada por uma plataforma *Online Judge* do Brasil. Os resultados obtidos nos apontam diferenças entre os modelos em relação às habilidades estimadas, diferenças que acreditamos estar relacionadas à forma com que cada modelo estima os parâmetros.

Palavras-chave: Habilidade, TRI, ELO.

ELO and TRI: Estimating Students Skill on a Programming Online Platform

Abstract. Methods of assessing student proficiency have gained prominence in recent years. There are a growing number of online courses and platforms that provide repositories of questions or exercises where assessment methods take place automatically. This paper analyzes the data generated by two models that aim to estimate students' ability: ELO and TRI. ELO is designed to rank players through game history while TRI estimates skill through a set of responses to a set of items. We use a database provided by an Online Judge do Brazil platform. The results show us differences between the models regarding the estimated abilities, differences that we believe are related to the way each model estimates the parameters.

Keywords: Skill, TRI, ELO.

1. Introdução

Melhorar os métodos de avaliação da proficiência e habilidades dos estudantes é um dos objetivos dos pesquisadores na área de educação. Há um número crescente de cursos e plataformas *online* voltados para o ensino de programação, que disponibilizam repositórios de exercícios onde os estudantes ou usuários submetem a solução dos mesmos e, muitas vezes, tem um *feedback* automático.

Os métodos de avaliação educacional com maior destaque são aqueles que buscam apresentar dados precisos sobre a construção das competências, sejam de estudantes presenciais ou à distância através do uso de plataformas *online* que utilizam avaliações automáticas. Isso ocorre porque há um grande número de usuários acessando essas

plataformas e, geralmente, os grupos são heterogêneos, isto é, existem usuários com bastante conhecimento de programação, outros nem tanto. Assim, o propósito deste trabalho é analisar dados obtidos a partir de dois métodos que estimam a habilidade dos indivíduos: Teoria de Resposta ao Item (TRI) e ELO.

Segundo Andrade (2000), a TRI é um conjunto de modelos matemáticos, que busca representar a probabilidade de um indivíduo acertar a resposta referente a um item em função dos parâmetros do mesmo e da(s) habilidade(s) do indivíduo. Tal relação, entre a probabilidade de um indivíduo dar uma certa resposta a um item e seus traços latentes, se expressa de forma que quanto maior a habilidade do indivíduo, maior será a probabilidade de acerto no item.

O ELO (Elo, 1978), originalmente criado para avaliar habilidades de jogadores de xadrez, objetiva classificar os jogadores através dos seus históricos de jogo. É um sistema de classificação estatística que pode calcular valores relativos ao nível de habilidade para competidores ou máquinas em jogos competitivos. Neste trabalho utilizamos um ELO adaptado (Prisco, 2018) que relaciona o estudante ao problema com o qual interage.

Em nossos experimentos, utilizamos um banco de dados de uma plataforma *Online Judge* do Brasil (Tonin, 2012), que contém problemas de programação onde os usuários (estudantes) da plataforma resolvem esses problemas e recebem um *feedback* de acerto ou erro. Esse experimento, com ELO e TRI, tem por finalidade entender as relações entre os dois modelos e descobrir, dentre os dois, o que melhor se aplica em um modelo de avaliação das habilidades de plataformas *Online Judge*.

A seguir, apresentaremos o referencial teórico que embasou nossa pesquisa, a metodologia utilizada para a execução dos experimentos e por fim, os resultados obtidos.

2. Referencial Teórico

Nesta seção são contextualizadas as abordagens de avaliação das habilidades dos estudantes e a plataforma virtual de aprendizagem.

2.1. Teoria de Resposta ao Item (TRI)

A TRI é utilizada para avaliação de medidas padronizadas e vem, gradativamente, sendo inserida no espaço da educação por ser considerada um instrumento importante no processo quantitativo de avaliação educacional, pois ela permite a construção de escalas de habilidades calibradas (Andrade, 2000) e (Baker, 2001). A TRI compreende modelos matemáticos para obter as estimativas das habilidades, também chamada de *Theta* θ , através das respostas dadas ao conjunto de itens em determinada área de conhecimento a ser avaliada. A estimativa da habilidade tem relação com a probabilidade de o indivíduo acertar o item, considerando um ou mais parâmetros. Dessa forma, quanto maior a habilidade do indivíduo, maior será a probabilidade de acerto ao item.

Para itens dicotômicos, há três modelos que diferem entre si pela quantidade de parâmetros utilizados para descrever o item. São eles: Modelo de Rasch (considera somente a dificuldade do item), Modelo de 2 parâmetros ou 2PL (considera a dificuldade do item e a discriminação do mesmo) e Modelo de 3 parâmetros ou 3PL (considera a dificuldade, a

discriminação e a probabilidade de acerto ao acaso) (Andrade, 2000) e (Baker, 2001). A Equação 1, descreve a probabilidade do indivíduo com habilidade θ acertar o item j , levando em consideração a discriminação do item a , a dificuldade do item b e a chance de acerto ao acaso c (Baker, 2001).

$$P(\theta) = c_j + (1 - c_j) \frac{1}{1 + e^{-a_j(\theta - b_j)}}$$

Equação 1: Modelo logístico de 3 parâmetros

O parâmetro de discriminação do item a indica o quanto um item distingue os indivíduos com diferentes níveis de habilidade. A dificuldade do item b se dá na mesma escala da habilidade, ou seja, a habilidade que o indivíduo tem que possuir para uma dada probabilidade de responder corretamente o item. O parâmetro de acerto ao acaso c é a probabilidade de um indivíduo com baixa habilidade acertar casualmente um item (Tavares, 2014) e (Oliveira, 2017). Quando não for permitido responder ao acaso, o parâmetro c assume valor 0 (zero) e o modelo que se aplica é o de 2 parâmetros (Andrade, 2000) e (Oliveira, 2017).

Na TRI, a habilidade tem relação com a probabilidade do indivíduo responder corretamente um item através de funções matemáticas. Tais funções precisam ser estimadas e são chamadas de "Curva Característica do Item" (CCI). A CCI representa a relação existente entre a habilidade e o desempenho nos itens. Mais informações sobre a CCI, pode ser encontrada em (Baker, 2001) e (Andrade, 2007).

Um dos principais passos da TRI é a estimação dos parâmetros dos itens (calibração) e das habilidades (traços latentes) e essa etapa necessita de recursos computacionais. Os modelos mais utilizados são o Método Bayesiano, que estabelece distribuições a priori para os parâmetros de interesse, constrói-se uma nova função denominada distribuição a posteriori e estima-se os parâmetros de interesse com base em alguma característica dessa distribuição (Andrade, 2000) e (Araujo, 2009).

Outro método utilizado é o de Máxima Verossimilhança Marginal que, através de iterações, estima os parâmetros (dos itens e habilidade) em duas etapas: na primeira, como o processo de estimação dos parâmetros começa a partir das respostas dos indivíduos, pois somente esses dados são conhecidos, estima-se os parâmetros dos itens, assumindo uma certa distribuição para as habilidades; na segunda, assumindo que os parâmetros dos itens já são conhecidos, estima-se as habilidades. Essa iteração ocorre até que já não haja mais variações significativas nas estimativas (Andrade, 2000), (Araujo, 2009) e (Chalmers, 2012).

Após a estimativa dos parâmetros dos itens, é preciso fazer uma análise desses resultados, pois valores críticos dos parâmetros implicam na exclusão do item para estimar as habilidades. Valores de discriminação abaixo de 0.30 não é considerado um item capaz de diferenciar indivíduos com diferentes estimativas de habilidade. Em relação ao índice de dificuldades, valores menores que -2.95 ou maiores que 2.95, também não são considerados itens adequados já que a escala de habilidade varia entre -3 e 3. O mesmo acontece com o acerto ao acaso, probabilidade maior de 0.40 também é considerado crítico. Nesses casos, se recomenda a eliminação dos itens (Andrade, 2000) e (Vendramini, 2005).

2.2. Sistema de Classificação ELO

O sistema de classificação ELO foi proposto para avaliar jogadores de xadrez. Cada jogador recebe uma classificação inicial (um ELO) e, à medida que vai participando dos jogos, essa classificação vai sendo atualizada de acordo com os resultados. O modelo trabalha em função da expectativa e do resultado. Se o resultado atende à expectativa (por exemplo, um jogador com maior habilidade vence um jogador com menor habilidade) os valores de ELO recebem pequenas atualizações, já em caso contrário, a atualização é maior (Elo, 1978) e (Pelánek, 2016). Em Prisco (2018) é apresentada uma adaptação da técnica ELO semelhante à usada na escolha de oponentes em torneios de xadrez ou partidas *online*. A expectativa de que i ganhe de j ($R_{ij} = 1$) é dado pela Equação 2 (Prisco, 2017):

$$P(R_{ij} = 1) = \frac{1}{1 + 10^{\frac{\theta_j - \theta_i}{400}}}$$

Equação 2: Modelo ELO

Onde $R = \{0,1\}$ é o conjunto de resultados de um jogo: 1 (ganhar) e 0 (perder), dado um jogo entre o jogador i e o jogador j , com um ELO θ_i e θ_j , respectivamente. Ao final da partida novos ELOs são calculados de acordo com as expectativas dos resultados, os ELOs anteriores e uma constante k . Quanto maior o k , maior é a mudança do ELO.

Nesse trabalho, para cada submissão consideramos um jogo, ou seja, um "duelo" entre o jogador e o problema. Em cada início de submissão, armazenamos o ELO do problema e do usuário ou estudante. Um *feedback* positivo (aceito) indica que o estudante "ganhou o jogo" enquanto que um *feedback* negativo (resposta errada, *time limit* e etc.) indica que o "problema ganhou". Assim, após a avaliação, os ELOs, do usuário e do problema, são atualizados.

O ELO pode ser utilizado na educação, se julgarmos que uma tentativa de solução tem uma relação entre item e aluno, estimando a habilidade do aluno e a dificuldade do item (Pelánek, 2016). A estimativa acontece de maneira contínua, pois a atualização da classificação acontece ao término de cada evento.

Segundo Pelánek (2016), existe uma relação entre o sistema de classificação ELO e a TRI de um parâmetro (modelo Rasch). O que os difere é o procedimento de estimativa dos parâmetros e em suas suposições básicas: TRI assume que a habilidade do estudante é constante, enquanto que o ELO foi implementado para rastrear as mudanças nos níveis de habilidades.

2.3. Online Judges

Online Judges são ambientes virtuais de aprendizagem que possuem um repositório de problemas de programação onde estudantes ou usuários podem competir resolvendo tais problemas (Tonin, 2012).

Professores e alunos incorporaram o uso dessa plataforma como uma ferramenta didática. Para resolver o problema, os estudantes devem elaborar um algoritmo (programa) em uma das linguagens de programação aceitas pela plataforma. A plataforma tem um sistema automático de avaliação que faz a análise do programa submetido e dá um *feedback* ao estudante. Os tipos de *feedback* são: aceito (problema aceito sem erro) ou de erro (compilação, execução, apresentação, tempo de execução, tempo excedido).

A plataforma *Online Judges* permite que o usuário resolva os problemas aleatoriamente e submeta uma quantidade indeterminada de resolução para o mesmo problema, mesmo que a resposta esteja correta.

Na base de dados disponibilizada, constam os *ids* dos usuários, *ids* dos problemas, data e hora que a solução foi submetida, a resposta da solução submetida (correta ou errada) e a linguagem de programação em que o problema foi resolvido. Em relação aos problemas as informações disponibilizadas são: *ID*, nome e nível.

3. Metodologia

Nesta seção, apresentamos a metodologia usada para aplicação do ELO e TRI. Utilizamos a base de dados disponibilizada pelo *Online Judge* no Brasil, com 62.997 usuários e 1.163 problemas distintos.

Como a TRI não leva em consideração a quantidade de submissões de cada problema, fizemos 2 experimentos: no primeiro, optamos somente pelas questões submetidas por todos os usuários, e elegemos somente os usuários que submeteram todas as questões selecionadas anteriormente. Para tabular a base, analisamos o conjunto de respostas de cada problema: mesmo submetendo mais de uma solução para o mesmo problema, se em alguma dessas a solução estivesse correta, consideramos a questão correta. Já no segundo experimento, a base de dados foi composta somente pelas questões que foram submetidas uma única vez.

Para ambos os experimentos foram eliminados os dados replicados e os itens em que todos usuários acertaram ou erraram, seguindo a sugestão da TRI de eliminação desses casos. Dessa forma, analisamos 1.006 usuários e 80 problemas (itens) no primeiro experimento e, no segundo experimento, foram analisados 3.357 usuários e 100 problemas. Os problemas corretos e que foram aceitos receberam o valor 1 (um), caso contrário, receberam o valor 0 (zero).

Na fase inicial, todos os usuários recebem um ELO (1100), de acordo com Prisco (2017). Para a atribuição de ELO dos problemas, usamos como heurística o nível de dificuldade indicado pelo autor do problema. Com os ELOs iniciais atribuídos, o algoritmo simula cada submissão realizada em ordem cronológica. Através do *feedback* de cada envio, os ELOs são atualizados. No final do processo, temos os ELOs finais de cada usuário e de cada problema. Os valores no final do processo são comparados com os valores pré-processados, se a diferença entre os valores é substancial isso indica que os ELOs ainda não convergiram e o processo é repetido. Enquanto o processo converge, nós armazenamos os ELOs finais de cada problema. Quando conhecemos o ELO de cada problema, consideramos que este seja constante para o próximo passo. Uma vez que a etapa de convergência tenha sido concluída, os usuários novamente recebem ELOs médios e, para cada nova submissão, seus ELOs são atualizados e armazenados no banco de dados. Dessa forma, temos o histórico do ELO de cada usuário para cada submissão.

Para aplicarmos a TRI, tabulamos as informações de forma a relacionar os usuários com as respostas dos problemas. A Tabela 1 mostra um exemplo dos dados tabulados, onde nas colunas constam os problemas e nas linhas os usuários.

Tabela 1: Dados tabulados para TRI

	1001	1002	1003	1004	1005	1006	1007	1008	1009	1010
1	1	0	1	0	1	1	1	0	1	1
2	1	1	1	0	1	1	1	0	1	1
3	0	1	1	1	0	0	1	0	0	0
4	1	0	1	1	1	0	1	1	0	0
5	0	1	1	1	1	1	0	1	1	1

Tendo em vista que a plataforma *Online Judge* não apresenta problemas de múltipla escolha, o modelo da TRI escolhido foi o de 2 parâmetros (2PL), modelo que não estima o acerto ao acaso, atribuindo o valor zero ao parâmetro c da Equação 1.

Para a aplicação da TRI, utilizamos a linguagem R e o pacote *mirt* que foi criado para estimar os parâmetros dos itens e a habilidade (traço latente), utilizando o método Máxima Verossimilhança Marginal (Chalmers, 2012). O primeiro passo foi fazer a calibração e parametrização dos itens para obter o nível de dificuldade e a discriminação dos mesmos.

Com os resultados dos parâmetros dos itens foi feita uma análise comparando-os com os valores considerados críticos, o que proporcionou a validação dos itens considerados satisfatórios, segundo as características da TRI, bem como a exclusão ou reavaliação dos itens considerados fora destas condições. Todos os itens analisados foram satisfatórios.

Após a estimativa dos parâmetros, o próximo passo foi obter o traço latente ou a habilidade dos usuários. De posse das habilidades encontradas através da TRI e ELO, realizamos a análise dos dados, discutidos na próxima seção.

4. Resultados e Discussões

Após a aplicação da metodologia e obtidos os resultados das habilidades nos dois modelos realizamos as análises através de dois experimentos.

No experimento 1, analisando o histórico de cada um dos usuários, foi possível perceber que o valor do ELO varia de acordo com as submissões, isso ocorre porque o ELO considera todas as tentativas de resolução dos problemas. Ou seja, aquele usuário que submete a solução do problema e a mesma está correta, tem seu ELO atualizado numa determinada escala, simultaneamente o ELO do problema também é atualizado, diminuindo seu valor ou a escala de dificuldade. O contrário também acontece, usuários que submetem várias vezes, até acertar a solução do mesmo problema, tem o seu valor de ELO atualizado, porém numa escala menor, conseqüentemente o ELO daquele problema aumenta, pois é considerado difícil pelo fato do usuário não ter acertado na primeira submissão.

A Figura 1 ilustra o histórico de submissões de 4 usuários, escolhidos aleatoriamente, com diferentes perfis de variação na habilidade ELO. O eixo horizontal apresenta a quantidade de submissões realizadas pelos usuários e o eixo vertical apresenta o valor de ELO a cada submissão.

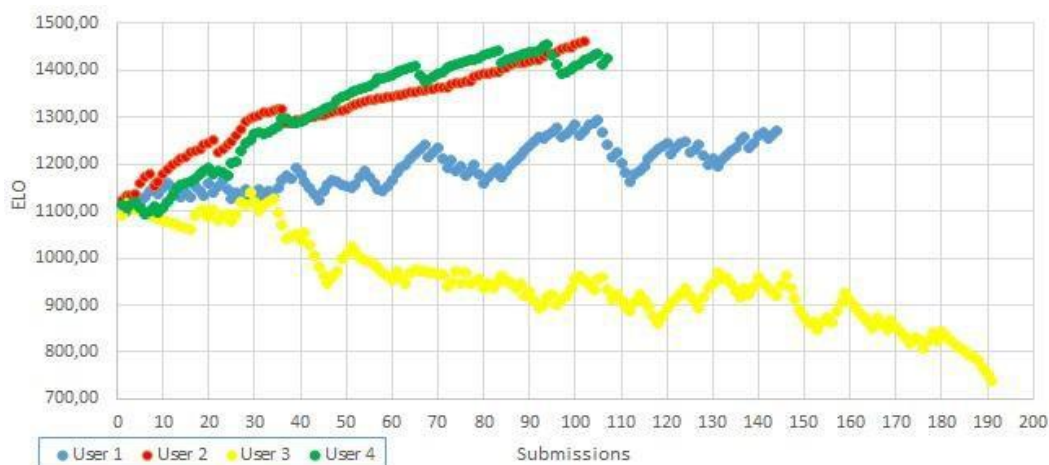


Figura 1: Histórico de submissões dos usuários - ELO

Tomaremos como exemplo o usuário 1, em azul, que submeteu 144 soluções, das quais 90 estavam corretas e 54 incorretas. Dos problemas corretos, 73 foram submetidos uma vez, 2 problemas foram submetidos 5 vezes e demais problemas foram submetidos entre 2 e 4 vezes, obtendo o acerto somente na última submissão. Um desses problemas com 5 submissões até o acerto, foi submetido entre a quadragésima primeira e a quadragésima quinta submissão, o que justifica a queda do valor ELO entre as escalas 41 e 45 no gráfico. O mesmo evento ocorre com os demais usuários e intervalos onde o gráfico ilustra o valor de ELO diminuindo e aumentando.

No geral, a maioria dos usuários ficou com ELO por volta de 1.200, ou seja, a maioria teve aumento do valor de ELO e os valores ELOS foram distintos, isto é, não houve ocorrência de mais de um usuário com o mesmo valor ELO. Porém, não houve uma dispersão muito grande entre os usuários devido ao nível de dificuldade dos problemas, isso se deu porque a maioria dos problemas analisados no experimento foram classificados no nível fácil pelos autores dos mesmos.

Na mesma base utilizada pelo ELO, foram estimados os níveis de dificuldade dos problemas e dos usuários pelo modelo TRI. Porém, para que pudéssemos comparar ELO com a TRI, selecionamos o ELO final de cada usuário, ou seja, o ELO da última submissão.

Na TRI foi considerado somente se o problema estava correto ou errado, independentemente do número de submissões até acertar o problema. Como a plataforma permite uma quantidade indeterminada de submissões, a maioria dos usuários submete várias soluções até obter acerto, logo essa quantidade de submissões foi desprezada. Dessa forma, o valor da habilidade acabou sendo o mesmo para muitos dos usuários. Para melhor analisar os resultados, foi feito um *ranking* de classificação das habilidades, o 1º ao 385º lugar. A Tabela 2 apresenta dados dos 3 primeiros *rankings*.

Tabela 2: *Ranking* TRI

<i>Ranking</i>	Quantidade de Usuários	Percentual de Usuários	Valor TRI
1º	510	50,69%	0,728
2º	55	5,4671%	0,586
3º	1	0,0994%	0,549

Dos 1.006 usuários, 50,69%, que corresponde a 510, obtiveram o maior valor de TRI 0,728, ocupando o primeiro lugar. Isso aconteceu porque todos eles acertaram todas as questões. O segundo lugar, está ocupado por 5,46% de usuários, que erraram a mesma questão, ficaram com TRI igual a 0,586. Apenas um usuário (0,099%) ficou em terceiro lugar no *ranking*, com o mesmo número de acertos dos usuários do *ranking* 2, porém com TRI no valor de 0,549.

Os usuários do *ranking* 2 e o usuário do *ranking* 3 erraram apenas uma questão, porém o parâmetro de discriminação dos problemas são distintos: o problema que os usuários do segundo *ranking* erraram é menor que o parâmetro de discriminação do problema que o usuário que está em terceiro lugar errou, por isso se deu a diferença no valor de TRI. O mesmo ocorreu com demais *rankings* e usuários, em que a dificuldade e as demais características das questões afetaram o resultado. Dessa forma a habilidade é estimada de acordo com o parâmetro de discriminação dos itens que os usuários acertam ou erram. Acertar a mesma quantidade de questões que outros usuários não significa que terão a mesma habilidade.

Na Tabela 3 relacionamos as habilidades dos dois modelos, ELO e TRI, dos mesmos usuários ilustrados na Figura 1. Tendo em vista que os valores dos ELOs variaram entre 555 e 1.475 e a TRI entre -3 e 0,50, a Tabela 3 mostra que nem sempre a relação dos valores das habilidades dos dois modelos é equivalente. Ou seja, o usuário com maior ou menor ELO pode não ser o mesmo com maior ou menor habilidade TRI. Por exemplo, observe os usuários 1 e 4, que apresentam ELOs altos e TRI baixos. Porém, se observarmos o usuário 2 percebemos habilidades altas nos dois modelos. Esse usuário é um dos que acertou todos os problemas, por isso recebeu o maior valor de TRI, mas não o maior valor de ELO, pois submeteu algumas soluções mais de uma vez, até acertá-las, o que fez diminuir o seu valor ELO.

Tabela 3: Habilidades estimadas pelo ELO e TRI

idUser	ELO	TRI
1	1272,18	-1,663830070
2	1461,77	0,49049927
3	740,088	-1,818720554
4	1424,5	-1,64809031

No experimento 2, aplicamos ELO e TRI na base dos problemas com apenas uma submissão. Nesse caso, da mesma forma que o ELO, a TRI já não apresentou ocorrência de usuários com a mesma habilidade, ou seja, cada usuário obteve um valor único de habilidade TRI, entre -3,52 e 1,97. No modelo ELO as habilidades variaram entre 384 e 3.003.

Nesse experimento fizemos a normalização das habilidades dos dois modelos para deixarmos na mesma escala. Após a normalização calculamos os percentis da diferença das habilidades, conforme ilustra a Figura 2.

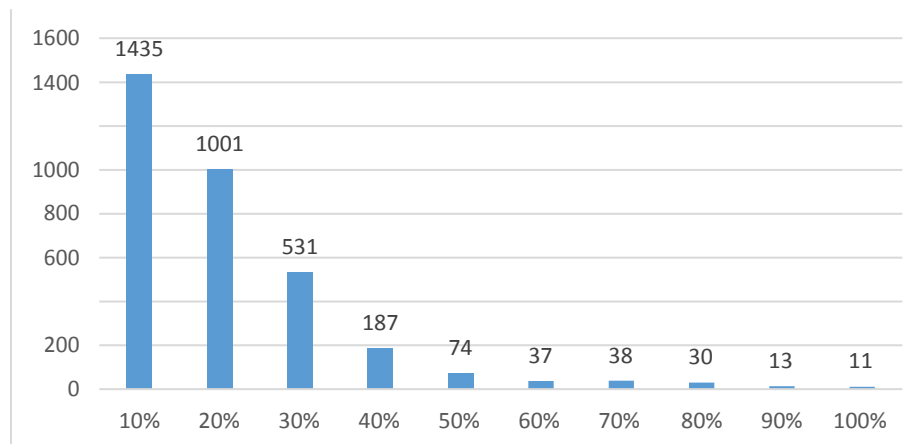


Figura 2: Percentis das habilidades

De acordo com o gráfico da Figura 2, dos 3.357 usuários, 1.435 estão na faixa de 10%. Isso significa que esses usuários têm 10% de diferença entre as habilidades ELO e TRI; 1.001 usuários estão na faixa dos 20% de diferença e, assim, sucessivamente. À medida que o percentual de diferença vai aumentando, a quantidade de usuários vai diminuindo, o que demonstra que não há grande diferença na estimativa das habilidades entre os modelos. Ainda podemos observar que 72,57% dos usuários apresentaram uma diferença de até 20% entre as habilidades, o que apoia a relação entre os modelos.

Na análise feita no experimento 1, nossa hipótese é que a diferença dos valores das habilidades possa estar relacionada à forma com que cada modelo estima os traços latentes. Os modelos da TRI supõem que a habilidade seja constante ou fixa, não leva em consideração o desempenho dos usuários ao longo do tempo, ou seja, no nosso experimento descarta o número de vezes que o usuário submeteu as soluções até acertar o problema. Já o modelo de classificação ELO acompanha a habilidade de mudança do usuário e adapta o grau de dificuldade dos problemas e habilidade dos mesmos, considerando todas interações entre o usuário e os problemas.

No segundo experimento, houve uma diferença considerada pequena entre os valores das habilidades entre os dois modelos, pois mais de 72% dos usuários obtiveram até 20% de diferença nos valores dos traços latentes. Nossa hipótese é que essa diferença diminuiu porque não foram consideradas as interações dos usuários com o problema, uma vez que cada problema teve apenas uma submissão por usuário. Nesse caso, os resultados mostram que ambos modelos apresentam estimativas mais próximas, corroborando com Pelánek (2016) que articula haver relação entre os dois modelos.

Como próximos passos planejamos pesquisar e entender modelos de TRI para itens não dicotômicos para estudar uma forma de estimar habilidades ao longo do tempo, unificando os dois modelos. Acreditamos que um modelo híbrido, que contemple ambas abordagens poderá ser aplicado na avaliação da habilidade e do processo de desenvolvimento de estudantes em plataformas *online* de programação.

5. Agradecimentos

Universidade Federal do Rio Grande, Instituto Federal de Educação, Ciência e

Tecnologia Sul-rio-grandense. O presente trabalho foi realizado, em parte, com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

6. Referências Bibliográficas

ANDRADE, D. F.; TAVARES, H. R.; VALLE, R. C. Teoria da resposta ao item: conceitos e aplicações. 2000. In: **SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA (SINAPE)**, São Paulo, 2000.

ANDRADE, D; JUSTINO, G. Software para avaliação de aprendizagem utilizando a teoria da resposta ao item. In **ANAIS DO XXVII CONGRESSO DA SBC**, Rio de Janeiro, 2007.

ARAÚJO, E. A. C.; ANDRADE, D. F.; BORTOLOTTI, S. L. V. Teoria de Resposta ao Item. **Revista da Escola de Enfermagem da USP**, São Paulo. 43(SPE):1000– 1008, 2009.

BAKER, F. B. **The basics of Item Response Theory**. ERIC, 2001. 187p.

CHALMERS, R. P. et al. mirt: A multidimensional Item Response Theory package for the R environment. **Journal of Statistical Software**, 48(6):1– 29, 2012.

ELO, A. E. **The rating of chessplayers, past and present**. Arco Pub, 1978. 220p.

OLIVEIRA, L. S. **O desempenho em matemática do ENEM de 2012 em Luis Eduardo Magalhães (BA), na Teoria de Resposta ao Item**. Arraias: UFT, 2017. 58p. Dissertação de Mestrado.

PELÁNEK, R. Applications of the Elo rating system in adaptive educational systems. **Computers & Education**, v. 98, 169–179, 2016. Disponível em: < <https://dl.acm.org/citation.cfm?id=2936090> >. Acesso em: 18 fev. 2019.

PRISCO, A., SANTOS, R., BOTELHO, TONIN, N., BEZ, J. Using information technology for personalizing the computer science teaching. In **2017 IEEE Frontiers in Education Conference (FIE)**, Indianapolis, 2017.

PRISCO, A., SANTOS, R., BOTELHO, S., TONIN, N., BEZ, J. A multidimensional Elo model for matching learning objects. In **2018 IEEE Frontiers in Education Conference (FIE)**, San Jose, 2018.

TAVARES, C. **A teoria de resposta ao item na avaliação em larga escala: Um estudo sobre o exame nacional de acesso ao mestrado profissional em matemática em rede nacional-profmat**. Rio de Janeiro: IMPA, 2014. 79p. Dissertação de Mestrado.

TONIN, N. A.; BEZ, J. L. URI Online Judge: A New Classroom Tool For Interactive Learning. In: **WORLDCOMP'12 - The 2012 World Congress in Computer Science, Computer Engineering, and Applied Computing**, Las Vegas. FECS 2012. USA: CSREA Press, 2012.

VENDRAMINI, C. M. M.; DIAS, A. S. Teoria de resposta ao item na análise de uma prova de estatística em universitários. **Psico-USF**, v. 10, n. 2. p.201–210, 2005.