

Uso da Escala de Stapel na Análise de Jogos

Bernardo Aguiar^a Walter Correia^b Fábio Campos^c

Universidade Federal de Pernambuco - UFPE

Resumo

Neste artigo, o terceiro numa série de trabalhos correlatos, iremos analisar como se comporta uma escala de autorrelato, a saber, a escala de Stapel, quando utilizada para avaliar jogos e aplicada através de um suporte online de fácil distribuição (a ferramenta questionário do Google.docs¹), de modo que possamos comparar o desempenho de tal escala com o desempenho de outras escalas de autorrelato, quando analisadas da mesma maneira, a fim de identificar seus pontos negativos e positivos.

Palavras-chave: Escalas de autorrelato, Stapel, avaliação de jogos

Contato dos autores:

a - Mestre em Design pela UFPE.
bcaguiar@hotmail.com

b - Doutor em Engenharia de Produção pela UFPE, professor e pesquisador da graduação e pós-graduação em Design da UFPE.
ergonomia@terra.com.br

c - Doutor em Ciências da Computação pela UFPE, professor e pesquisador da graduação e pós-graduação em Design da UFPE.
ffcc@ieee.org

1. Introdução

O uso de escalas de autorrelato (escalas onde o respondente avalia algo – um jogo, por exemplo – por si próprio, sem influência externa, através de respostas dadas num questionário) para coletar opiniões e avaliações objetivas é atividade comum e difundida entre várias áreas do saber.

Em Game Design, tais escalas são uma ferramenta ágil na coleta de avaliações sobre o jogo, seja durante

seu processo de produção, coletando avaliações da equipe como forma de identificar pontos a serem melhorados ou corrigidos; seja após o jogo ter sido lançado para o público, como forma de analisar a receptividade do mesmo frente aos jogadores.

Entre uma das muitas escalas de autorrelato existentes está a escala de Stapel. A escala de Stapel foi baseada na escala de diferencial semântico, com 11 níveis de gradação ao invés de sete e algumas outras diferenças na forma como se apresenta ao respondente.

Mas, no entanto, para melhor usar essas escalas, é preciso conhecer como elas se comportam quando postas em prática. Para tanto, no presente trabalho, analisaremos a utilização da escala de diferencial semântico na avaliação de jogos, através de experimentos onde se pediu a membros do capítulo Recife da International Game Developers Association - IGDA (através de sua lista de e-mails) e a gamers voluntários (representando o público alvo) para avaliarem, usando uma escala Likert desenvolvida especialmente para este experimento quatro jogos em Flash encontrados na internet e escolhidos pelo autor.

Nas próximas seções delinearemos quais parâmetros precisam ser seguidos para construir uma escala de diferencial semântico, como foram realizados os experimentos, como analisamos os resultados e quais conclusões foram retiradas do processo.

2. Trabalhos relacionados

Sendo o segundo numa série de artigos sobre o uso das escalas de autorrelato na avaliação de jogos, o presente artigo é diretamente relacionado aos outros dois: “Uso da Escala Likert na Análise de Jogos” e “Uso da Escala de Stapel na Análise de Jogos”, ambos pelos mesmo autores.

3. Escalas de Autorrelato

Escalas de autorrelato são usadas nos projetos de Design como formas de mesurar o desempenho de um produto frente ao seu público-alvo ou um conceito frente à equipe desenvolvedora para medir sua viabilidade. Servem também como formas de apurar a opinião das pessoas frente ao produto ou projeto (sendo, portanto, consideradas também como escalas psicométricas, quando usadas desta maneira).

Empresas podem se utilizar da aplicação de tais escalas para pesquisar seu público-alvo e mensurar a receptividade que seus produtos apresentam frente aos usuários. Na indústria de jogos casuais para web

¹ O Google.docs é uma plataforma online de aplicativos estilo Office, com alternativas de uso gratuito e com a possibilidade de acesso colaborativo entre vários autores. Entre os aplicativos esperados de uma suíte Office (processador de textos, planilhas de cálculo e construtor de slides), existe também uma ferramenta para a elaboração e aplicação de questionários online.

encontramos uma facilidade maior de aplicar tais escalas: utilizando a internet como plataforma de suporte para as escalas, pode-se contar com a familiaridade prévia dos jogadores com o ambiente online para responderem os questionários, além de se tornar mais fácil levar os questionários para o público quando utilizamos a internet como suporte.

De posse desses dados, é possível então realizar ajustes aos jogos já feitos ou evitar tendências que alienem o usuário em jogos futuros e, com a aplicação via internet dos formulários com as escalas de autorrelato, diminui-se o tempo e mobilização necessários para realização da pesquisa de opinião, podendo assim ter retornos mais imediatos e um eventual reuso dos formulários, adaptando-os para colher informações sobre jogos futuros ou ainda sobre outros aspectos de jogos já produzidos.

As escalas de autorrelato se tornam ferramentas poderosas, então, auxiliando, complementando e validando as informações disponíveis sobre o usuário final. Servindo como suporte durante várias fases do projeto, as escalas são uma forma ágil para coletar opiniões e simples de serem aplicadas, podendo ser incluídas sem causar um distúrbio grande no projeto.

4. Escalas de Stapel

Uma variante da escala de diferencial semântico, consistindo em uma escala verbal de 11 pontos, com valores de -5 a +5, com o termo a ser avaliado na posição do meio (ou zero), representando uma posição neutra em relação ao item.

Os entrevistados avaliam quão precisamente cada termo descreve o produto avaliado marcando uma dessas opções em cada um dos itens. Segundo Kinnear e Taylor [1991], “essa técnica de escalonamento é designada para mensurar a direção e a intensidade de atitudes, simultaneamente”.

Churchill [1998, *apud* Oliveira, 2001] diferencia as escalas de Stapel das escalas de diferencial semântico através das seguintes características: Os adjetivos ou frases descritivas são testados separadamente, ao invés de simultaneamente; os pontos na escala são identificados por números; e há onze pontos na escala, e não sete.

O fato de não precisarem ser definidos adjetivos ou frases descritivas bipolares é um dos pontos positivos desse tipo de escala, facilitando sua elaboração. A análise de seus resultados se dá através do cálculo das médias dos valores de suas respostas.

5. Experimento

O experimento, semelhante ao usado nos outros artigos desta série (ver Item 2 – trabalhos relacionados) consiste na aplicação de uma escala de Stapel de 11

gradações elaborada online para a avaliação de quatro jogos em Flash, escolhidos dos portais *Newgrounds.com* e *Nitrome.com*; e enviada para os integrantes da lista de discussões do capítulo Recife da IGDA (*International Game Developers Association* - Associação Internacional de Desenvolvedores de Jogos), representando o público profissional; e para voluntários interessados por jogos, representando o público-alvo.

O experimento estendeu-se por cinco semanas, entre novembro e dezembro de 2010. Em Cada uma das semanas os respondentes eram orientados a jogar um dos jogos e responder um questionário Stapel acerca dele. A quinta semana consistiu em uma semana extra para coleta de respostas adicionais.

Além do questionário de Stapel, os respondentes tinham como opção preencher um campo adicional, onde poderiam escrever suas considerações acerca do questionário em si, dando um retorno de sua experiência ao utilizá-lo.

No fim do processo, computamos as respostas da maneira descrita na seção 4 e montamos gráficos comparativos. Calculamos também o desvio-padrão das respostas de cada jogo, considerando uma distribuição normal, para avaliar a variação de respostas de cada item entre um respondente e outro.

O desvio-padrão é um “indicador que mostra a diferença da média de cada valor em um dado conjunto” [Takahashi, 2009]. Ele parte de 0 (o que indicaria que não houve variação nenhuma nas respostas de cada um) e, quanto mais alto seu valor, mais respostas distantes da média houve. Desse modo podemos usar o desvio-padrão como um indicador da convergência ou divergência das respostas dadas.

Por fim, calculamos a média dos desvios-padrões obtidos em cada uma das três categorias como forma de obter uma visão geral do desempenho da escala, a qual dividimos em três categorias: Gráficos e Visual; Sons e Música; e Jogabilidade e Mecânicas de jogo, com seis itens cada².

5.1. Resultados do jogo *3 Foot Ninja II*

[<http://www.newgrounds.com/portal/view/192451>]

Com 14 respondentes, o jogo *3 Foot Ninja II* obteve resultados muito negativos quando avaliado.

A média de seus desvios-padrão foi de 2,80. Quando comparado em proporção ao total de 11

² As tabelas com os desvios-padrão de cada jogo podem ser encontradas no item 8. Gráficos mais detalhados com as respostas médias de cada item podem ser encontrados em: [<http://bit.ly/npxqAn>]

opções de resposta em cada item, equivale a 25,45% das possibilidades.

5.2. Resultados do jogo *Faultline*

[<http://www.nitrome.com/games/faultline>]

Dos quatro jogos avaliados, este foi o que obteve mais respondentes, com 19 participantes respondendo. Os desvios-padrão ficaram entre 1,13 e 3,37, com um valor médio de 2,37.

5.3. Resultados do jogo *Brink of Alienation*

[<http://www.newgrounds.com/portal/view/254901>]

Com um total de apenas 10 respondentes, esse foi o jogo com menos respostas (um reflexo talvez das opiniões quanto a ele, em geral bastante negativas).

Os resultados obtidos foram ruins, com opiniões negativas predominando. Os desvios-padrão ficaram entre 0,79 e 3,01, com uma média de 2,09.

5.4. Resultados do jogo *Final Ninja Zero*

[<http://www.nitrome.com/games/finalninjazero>]

Com 17 participantes respondendo seus questionários, foi também o segundo a obter mais avaliações. Os desvios-padrão ficaram entre 1,94 e 3,13, possuindo um valor médio de 2,57.

5.5. Impressões dos participantes

Como parte dos experimentos, disponibilizamos uma caixa de texto, após cada questionário, para que os participantes que assim o desejassem dessem seu retorno acerca das escalas.

A escala foi recebida de forma bastante negativa: Seja pela falta de familiaridade com a mesma; seja pelo fato da opção central, mesmo indicando a opção neutra, tender a criar um juízo prévio (já que o próprio termo selecionado possui uma carga de significado difícil de desprender); sendo avaliada como confusa e pouco intuitiva.

6. Considerações finais

Apesar da simplicidade e facilidade de aplicação e distribuição da escala de Stapel através de um meio online, os níveis numéricos que servem como abstração para os graus de concordância ou discordância em um dado item causaram incômodo nos participantes, que não se sentiam seguros ao marcar certa opção como a que corresponde à sua opinião.

Além disso, a escala obteve altos níveis de desvios-padrão. Como as medidas do desvio-padrão são dadas na mesma unidade dos valores usados em seu cálculo e os valores médios deste ficaram entre 2,09 e 2,80, o que nos permite afirmar que, considerando uma distribuição normal, aproximadamente dois terços dos usuários tiveram respostas variando entre a resposta média e dois graus acima ou abaixo (chegando perto dos três graus de diferença).

Essa discrepância entre a resposta média e a distribuição das respostas de cada usuário causa problemas em identificar os pontos positivos e negativos do jogo através da utilização da escala de Stapel, principalmente quando o resultado médio fica em torno do ponto neutro, onde o desvio-padrão mostra a possibilidade da opinião dos respondentes estarem pendendo para um lado ou outro do espectro positivo/negativo de avaliação.

Apesar disso, a facilidade de elaboração e distribuição através do Google.docs foi um ponto positivo, agilizando a aplicação e coleta dos resultados dos questionários, além de liberar os respondentes da necessidade de estarem presentes fisicamente no ato de preencherem suas respostas.

7. Conclusão e desdobramentos futuros

Ao fim do processo, podemos concluir que o uso da escala de Stapel na coleta de opiniões durante o processo de produção de jogos, apesar de poder se tornar uma ferramenta poderosa para a aquisição de uma visão geral das impressões dos jogadores, possui uma série de características que denigrem sua utilidade quando usada desta maneira.

Em trabalhos futuros compararemos os resultados obtidos pela escala de Stapel (avaliada no presente artigo), Likert e diferencial Semântico (avaliadas em outros dois trabalhos dos mesmos autores), em busca de quais pontos positivos de cada uma delas podem ser usados na formulação de uma escala que funcione como uma alternativa, procurando potencializar os pontos benéficos e minimizar os problemas percebidos pelos usuários.

8. Tabelas com os desvios-padrão

Desvio-Padrão	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Gráficos e visual	2,45	2,66	2,73	2,19	2,88	3,58
Sons e música	2,52	2,57	2,63	1,82	2,11	2,73
Jogabilidade e mecânicas de jogo	3,95	3,13	3,84	2,73	3,00	2,54
Média dos desvios-padrão	2,80					

Tabela 1: Desvios-padrão e média dos desvios-padrão dos resultados das respostas do jogo *3 Foot Ninja II*.

Desvio-Padrão	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Gráficos e visual	2,33	2,34	2,57	3,37	2,04	2,85
Sons e música	2,57	2,95	2,80	2,14	2,05	1,71
Jogabilidade e mecânicas de jogo	2,21	1,74	2,30	2,67	1,13	2,91
Média dos desvios-padrão	2,37					

Tabela 2: Desvios-padrão e média dos desvios-padrão dos resultados das respostas do jogo *Faultline*

Desvio-Padrão	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Gráficos e visual	0,79	2,00	2,16	2,91	0,79	2,86
Sons e música	1,70	2,32	1,57	2,32	2,76	2,58
Jogabilidade e mecânicas de jogo	1,16	1,16	2,32	2,30	3,01	3,01
Média dos desvios-padrão	2,09					

Tabela 3: Desvios-padrão e média dos desvios-padrão dos resultados das respostas do jogo *Brink of Alienation*

Desvio-Padrão	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
Gráficos e visual	2,06	2,42	2,82	2,28	1,94	2,08
Sons e música	2,88	3,13	2,45	2,26	2,37	3,04
Jogabilidade e mecânicas de jogo	2,42	2,62	3,09	2,64	2,76	3,04
Média dos desvios-padrão	2,57					

Tabela 4: Desvios-padrão e média dos desvios-padrão dos resultados das respostas do jogo *Final Ninja Zero*

Agradecimentos

Os autores gostariam de agradecer a todos aqueles que participaram dos experimentos e que contribuíram, de uma forma ou outra, ao desenvolvimento do artigo.

Referências

- BLASTERMASTER, 2005. Brink of Alienation. [Jogo online]. Disponível em: <http://www.newgrounds.com/portal/view/254901> [Último acesso: 29 jul 2011]
- CHISNALL, P., 1973 Marketing Research: Analysis and measurement. Nova Iorque: McGraw Hill
- CHURCHILL, G., 1998. Marketing research: methodological foundations. 3a edição. Nova Iorque: The Dryden Press,
- FERREIRA, R., 2010. Performance das Técnicas de Geração e sua Medição. Dissertação de Mestrado. Recife: Universidade Federal de Pernambuco
- KINNEAR, T. C.; TAYLOR, J. R., 1991. Marketing Research: An applied approach. Nova Iorque: McGraw Hill
- MATTAR, F., 1996. Pesquisa de Marketing, vol. 1. São Paulo: Atlas
- MINICLIP, 2004. 3 Foot Ninja II. [Jogo online]. Disponível em: <http://www.newgrounds.com/portal/view/192451> [Último acesso: 29 jul 2011]
- NITROME, 2010. Faultline. [Jogo digital]. Disponível em: <http://www.nitrome.com/games/faultline/> [Último acesso: 29 jul 2011]
- _____, 2009. Final Ninja Zero. J[Jogo digital]. Disponível em: <http://www.nitrome.com/games/finalninjazero/> [Último acesso: 29 jul 2011]
- OLIVEIRA, B., 2010. Cards Persona: Aplicação da técnica de personas na criação de jogos digitais. Dissertação de Mestrado. Recife: Universidade Federal de Pernambuco
- OLIVEIRA, T., 2001. [online] Escalas de Mensuração de Atitudes: Thurstone, Osgood, Stapel, Likert, Guttman, Alpert. Disponível em: http://www.fecap.br/adm_online/art22/tania.htm [Último acesso: 30 jul 2011].
- PARASURAMAN, A., 1991. Marketing Research. Nova Iorque: Addison-Wesley publishing co.
- TAKAHASHI, S., 2010. Guia Mangá de Estatística. São Paulo: Novatec Editora
- TEIXEIRA, G., 2005. Conheça o que são Escalas de Medida. [online] Disponível em: <http://www.serprofessoruniversitario.pro.br/ler.php?modulo=21&texto=1304> [Último acesso: 10 mai 2011]