



Soraia Teixeira Pires

**Avaliação Psicológica via internet: um estudo de validade e fidelidade das versões
online (versus papel-lápis) de uma medida de sintomas depressivos**

Universidade Lusófona do Porto
Faculdade de Psicologia, Educação e Desporto

Porto
Outubro, 2017



Soraia Teixeira Pires

Avaliação Psicológica via internet: um estudo de validade e fidelidade das versões online (versus papel-lápis) de uma medida de sintomas depressivos

Dissertação defendida em provas públicas na Universidade Lusófona do Porto no dia 03 de Janeiro de 2018, perante o seguinte júri:

Presidente: Prof^ª. Doutora Inês Martins Jongenelen (Prof^ª. Auxiliar da Universidade Lusófona do Porto)

Arguente: Prof^ª. Doutora Cátia Margarida dos Santos Pereira de Oliveira (Prof^ª. Auxiliar da Universidade Lusófona do Porto)

Orientador: Prof. Doutor Diogo Jorge Pereira Lamela (Prof. Auxiliar da Universidade Lusófona do Porto).

Porto

Outubro, 2017

Não é autorizada, por um prazo de 3 anos, a reprodução de qualquer parte desta tese/dissertação.

Índice

Agradecimentos	7
Abstract	8
Resumo	9
Introdução	10-13
Método	13
Participantes	13-14
Medidas	14-15
Procedimentos	15-16
Análise Estatística	16
Resultados	16-20
Discussão	20-23
Referências	23-26

Índice de tabelas

Tabela 1- <i>Análise das frequências dos itens do PHQ9 respondido online</i>	17
Tabela 2- <i>Análise das frequências dos itens do PHQ9 respondido papel-e-lápis</i>	18
Tabela 3- <i>Coeficientes de saturação estandardizados dos itens da solução fatorial do PHQ9 para amostra online</i>	19
Tabela 4- <i>Coeficientes de saturação estandardizados dos itens da solução fatorial do PHQ9 para amostra papel-e-lápis</i>	19
Tabela 5- <i>Diferenças entre as amostras nos itens e total do PHQ9</i>	20

Lista de Abreviaturas

AFC – Análise Fatorial Confirmatória;

BDI-II – *Beck Depression Inventory-II*;

EDS- *Ehlers-Danlos syndromes*

DP – Desvio-padrão

PHQ-9 – *Patient Health Questionnaire-9*;

Agradecimentos

Gostaria de agradecer ao meu Orientador, Professor Diogo Lamela, pela dedicação, disponibilidade e atenção prestada durante este processo. Agradecer por me ensinar todos os seus conhecimentos e ajudar-me a aumentar os meus conhecimentos, através da excelente capacidade de trabalho da sua parte. Sem a sua disponibilidade e dedicação não seria possível chegar ao fim desta etapa.

Gostaria de agradecer aos meus pais, Francisco Pires e Maria Pires, pelo apoio que me deram ao longo desta etapa, bem como tudo o que fizeram por mim, pois sem eles não seria possível concluir este meu objetivo.

Gostaria de agradecer ao Cristiano Ribeiro e ao Xavier Pires pela paciência que tiver comigo quando mais precisava da ajuda deles. Pelo empenho e dedicação que demonstram para comigo quando eu precisava deles. A força que me deram foi muito importante para conseguir este meu objetivo.

Gostaria de agradecer a todos os participantes no estudo, pois sem a sua participação e colaboração não seria possível a realização deste estudo.

Obrigado a todos os que me apoiaram e ajudaram a concretizar um sonho da minha vida. O chegar ao fim de uma etapa que considero importante na minha vida académica.

Abstract

The present study aims to evaluate and compare the psychometric characteristics of a measure of depressive symptoms in its online version and pencil paper. Specifically, it aimed to test the sensitivity of the samples, to examine the construct validity of the PHQ9, to analyze the reliability of the PHQ9 and to test the differences in the averages of the items and the total scores of the PHQ9. Method: The study included 470 participants, of whom 278 were sample participants for the online version and 192 participants for the paper-pencil sample. The instrument administered was the Patient Health Questionnaire (PHQ-9). RESULTS: These data reveal that no statistically significant differences were found in the two samples, as well as in the total PHQ9 score. The construct's validity and reliability showed positive results, indicating that PHQ9 had good reliability and was statistically suitable for performing AFE in both the online sample and pencil paper. Conclusion: The PHQ9 can be used in the online version and the pencil-paper version without significant changes.

Keywords: Psychological Evaluation; Online version; Paper-pencil version.

Resumo

O presente estudo tem como objetivo avaliar e comparar as características psicométricas de uma medida de sintomas depressivos na sua versão *online* e papel-lápis. Concretamente, teve como objetivos testar a sensibilidade das amostras, examinar a validade de construto do PHQ9, analisar a fiabilidade do PHQ9 e testar as diferenças nas médias dos itens e dos scores totais do PHQ9. **Método:** O estudo incluiu 470 participantes, dos quais 278 eram participantes na amostra para a versão *online* e 192 participantes para a amostra papel-lápis. O instrumento administrado foi o Patient Health Questionnaire (PHQ-9). **Resultados:** Estes revelam que não foram encontradas diferenças estatisticamente nas duas amostras, bem como no score total do PHQ9. A validade do construto e a fiabilidade revelaram resultados positivos, indicando que o PHQ9 tinha uma boa fiabilidade e era estatisticamente adequado para a realização da AFE quer na amostra *online*, quer na papel-lápis. **Conclusão:** O PHQ9 pode ser utilizado na versão *online* e na versão papel-lápis sem que haja alterações significativas.

Palavras-chaves: Avaliação Psicológica; Versão *online*; Versão Papel-lápis.

A tecnologia tem evoluído ao longo dos anos, e a par dessa evolução surge o conceito de avaliação psicológica via *online*, uma prática que começa a ser utilizada por vários profissionais especializados (Carlbring, Brunt, Bohman, Austin, Richards, Goran & Andersson, 2007). É uma prática psicológica relativamente recente e tem sido alvo de vários estudos com objetivo de verificar a sua validade psicométrica comparativamente com a versão papel-lápis. A avaliação psicológica via *online* consiste em que os sujeitos respondam aos instrumentos psicológicos num local com acesso à internet, sem ter que se deslocar a um local específico. A versão *online* de instrumentos psicológicos pode apresentar vantagens metodológicas e clínicas. Em primeiro, permite ter mais facilidade no acesso, o que cria maior contacto com as populações mais isoladas, a fim de estas poderem ter acesso a esta prática, e tem vantagens ao nível do tempo, ou seja, é uma forma de os sujeitos responderem aos questionários em menor tempo, e tornando-se menos dispendiosa esta utilização (Carlbring et al., 2007).

Sheehan e McMillan (1999), num estudo sobre as variações das respostas via *online*, concluíram que o uso da internet para a avaliação psicológica tem vantagens ao nível económico, visto que não é necessário a utilização de material como, papel, tinteiros, correspondência.

A investigação indica ainda que não existem só vantagens na avaliação psicológica via *online* (Buchanan, 2003; Sheehan e McMillan, 1999). Há estudos que identificaram desvantagens no uso da avaliação psicológica via *online* comparativamente com a versão papel-lápis, pelo facto da versão online depender por um lado da velocidade da internet, e por outro, da conexão, ou seja, o questionário pode estar na fase final e a conexão falhar, tem de voltar a responder, e ao voltar a fazer o mesmo procedimento, os resultados não serão os mesmos, induzindo a outro tipo de respostas, e o resultado será outro (Buchanan, 2003). O facto de muitas pessoas não ter acesso às novas tecnologias (computador) e não ter acesso à internet, devido a dificuldades financeiras e/ou por não saber trabalhar com as mesmas, torna-se numa desvantagem para este tipo de avaliação (Buchanan, 2003). Contudo, as desvantagens mais salientadas nos estudos, referem-se às alterações psicométricas que os questionários sofrem ao serem transferidos para o formato *online* (Buchanan, 2003). Este autor realizou vários estudos nos quais investigava as propriedades psicométricas de vários questionários, a fim de identificar as alterações psicométricas que os questionários sofriam na sua versão *online*, sugerindo diferenças

assinaláveis nas qualidades psicométricas entre as respostas papel-lápis e *online*, em que método *online* pode criar riscos adicionais à validade e à fiabilidade das medidas. Já anteriormente, Woolhouse e Myers (1999), que compararam a aplicação de instrumentos *online* e de instrumentos em versão papel-lápis relativamente à personalidade, concluíram que haviam diferenças nas estruturas psicométricas dos instrumentos utilizados. Dessa forma, não se deve considerar como garantidas as propriedades psicométricas de um instrumento transferido para a versão *online*, mesmo que este seja transferido como o original da versão papel-lápis válido.

A confidencialidade e o anonimato das respostas são aspetos que se devem ter em atenção, uma vez que os mesmos podem estar em causa. Todavia, esse aspeto foi salientado e estudado em alguns estudos, e foi comprovado que não existe qualquer forma de ser violada a confidencialidade dos sujeitos, assim como o anonimato das respostas (Herrero & Meneses, 2006). Sendo, que após o processo estar concluído, são realizados todos os procedimentos de limpeza de dados, para que não existam qualquer tipo de dados que possam identificar o participante.

No entanto, a tendência da investigação mais recente revela não existir um comprometimento da qualidade psicométrica das medidas, em função do método de recolha de dados (papel-lápis vs. *online*), sendo estes resultados transversais a diferentes construtos psicológicos (Vallejo, Jordán, Díaz, Comeche & Ortega, 2007; Fouladi, McCarthy & Moller, 2002; Wood, Nosko, Desmarais, Ross & Irvine, 2006; Ward, Clark, Zabriskle & Morris, 2014; Bjornsdotter, Enebrink & Ghaderi, 2013). A título de exemplo, vários estudos (Coyne, 1999; Burke, Burke, & Baker, 1995; Ryan, Corry, & Attewell, 2002), que compararam a versão *online* versus papel-lápis do SF-36 (uma medida de percepção de saúde), encontraram resultados semelhantes entre ambos, o que significa que tem a mesma validade psicométrica. Adicionalmente, outros estudos que comparavam as duas versões em relação a outras medidas de personalidade, por exemplo *Eysenck Personality Inventory*, e a maior parte dos estudos reconheceram que ambas as versões se mantinham com o mesmo formato (Johnson, 2005; Cronk & West, 2002; Buchanan, Johnson & Goldberg, 2005).

Alguns sugerem, no entanto, que a semelhança das qualidades psicométricas dos instrumentos em função do tipo de recolha de dados pode não ser transversal a todos os construtos psicológicos, devido à natureza (e.g., personalidade, atitudes, estados afetivos) e base concetual (Buchanan, 2003). No que concerne aos instrumentos de estados afetivos a investigação quanto às qualidades psicométricas dos instrumentos papel-lápis e *online* é

ainda pouco sistematizada. Poucos estudos têm-se focado na investigação se o tipo de recolha de dados influencia a validade e a fiabilidade das respostas. No entanto, tem havido consenso empírico entre os estudos que as respostas *online* a instrumentos de avaliação dos sintomas depressivos não diferem das respostas papel-lápis na validade de construto, validade convergente, fiabilidade e sensibilidade e não estão sujeitas a maior viés ou a menor desejabilidade social (Herrero & Meneses, 2006; Vallejo et al., 2007; Zlomke, 2009; Spek, Nyklícek, Cuijpers & Pop, 2008; Hollandare, Andersson & Engstrom, 2010). Por exemplo, as conclusões que os estudos anteriormente descritos referem é que os instrumentos não sofrem alterações psicométricas, bem como podem ser utilizados na versão *online* todas as medidas psicométricas (Herrero & Meneses, 2006; Vallejo et al., 2007; Zlomke, 2009; Spek, Nyklícek, Cuijpers & Pop, 2008; Hollandare, Andersson & Engstrom, 2010). No entanto o estudo de Vallejo et al., (2007) refere que o instrumento SCL-90-R deve ser usado com algumas recomendações.

O presente estudo

Tendo em conta o movimento *e-mental health* que preconiza a utilização de meios de comunicação eletrónica na prestação de serviços psicológicos, surge como relevante compreender se as medidas de avaliação psicológicas são validadas previamente em papel-lápis apresentam comportamento psicométrico similar na sua versão *online* (Riva, Teruzzi & Anolli, 2004). Esta examinação prévia torna-se urgente na medida em que deve ser assegurada que as intervenções *online* são sustentadas em avaliações psicológicas *online* fiáveis, válidas e que cumpram os critérios psicométricos das medidas psicológicas (Nguen, Klein, Meyer, Austin & Abbott, 2015). Todavia, do nosso conhecimento, nenhum estudo foi desenhado em Portugal até ao momento para analisar o comportamento psicométrico de um instrumento psicológico na sua versão *online* e papel-lápis. Num contributo inicial, o presente estudo tem como objetivo avaliar e comparar as características psicométricas de uma medida de sintomas depressivos na sua versão *online* e papel-lápis. Dois motivos orientaram a seleção de uma medida de sintomas depressivos para testar o nosso objetivo geral. Em primeiro, alguns estudos sugeriram que são os instrumentos que avaliam os estados afetivos aqueles que podem apresentar maior variabilidade do comportamento psicométrico em função do tipo de recolha de dados (Donker, Straten, Marks & Cuijpers, 2010), pelo que maior atenção deve ser dada a construtos psicológicos. Em segundo, os sintomas depressivos são um dos grupos sintomáticos com maior prevalência na população (Kessler & Bromet, 2013), pelo que são

alvo de maior intervenção e *screening*, o que torna clinicamente relevante a compreensão da adequação destas medidas para uso *online*. O instrumento-alvo deste estudo será o Patient Health Questionnaire-9 (Kroenke, Spitzer, & Williams, 2001). Atualmente, este instrumento é uma das medidas mais internacionalmente utilizadas para despiste de sintomas depressivos, dada a sua brevidade (9 itens), validade, sensibilidade e poder discriminativo entre pessoas deprimidas e não deprimidas. Apesar de estudos comparativos entre versões *online* e papel-lápis de medidas de humor depressivos tem sido desenvolvido na literatura internacional, nenhum estudo similar foi conduzido com esta medida, apesar de ser uma das mais utilizadas em contextos de investigação e de cuidados primários.

Assim, este estudo teve quatro objetivos específicos. O primeiro objetivo foi testar a sensibilidade das respostas ao instrumento nas duas amostras. Colocou-se a hipótese que a sensibilidade dos resultados seria demonstrada em ambas as amostras, avaliada através da representação de todas as categorias de resposta da escala de *Likert* para cada um dos itens do PHQ-9.

O segundo objetivo foi examinar a validade de construto do PHQ-9 em cada um dos métodos de recolha de dados, testando a estrutura fatorial da medida em cada uma das amostras através de uma análise fatorial exploratória (AFE). Hipotetizou-se que não haveria diferenças significativas nos indicadores de validade de construto, em função do método de recolha de dados. O terceiro objetivo foi analisar a fiabilidade do PHQ-9 em função do método de recolha de dados. Esperou-se que os valores de alfa *Cronbach* fossem similares entre as participantes que responderam ao PHQ-9 *online* e as participantes que responderam em papel e lápis. Finalmente, o quarto objetivo do estudo foi testar diferenças nas médias dos itens e dos *scores* totais no PHQ-9, em função do método de recolha de dados. Como ambas as amostras eram constituídas por participantes da comunidade e, presumivelmente, com baixo risco de elevados sintomas depressivos, colocou-se a hipótese que não haveria diferenças nos *scores* em função do método de recolha de dados.

Método

Participantes

A amostra total foi constituída por 470 participantes do sexo feminino, que o foi resultado congregado de duas subamostras: uma amostra de participantes que respondeu *online* e uma amostra que respondeu papel-lápis.

Relativamente às características sociodemográficas da amostra total, a idade média foi de 35.63 anos ($DP = 9.73$ anos). Ao nível do estado civil, 66.2% eram casados ou união de facto, 26.8% solteiros, 6.0% divorciados ou separados e 0.4% viúvos. Aproximadamente, 13.4% dos participantes relataram terem uma escolaridade igual ou inferior ao nono ano, 30.6% detinham ensino secundário e 54.5% tinham um grau obtido no ensino superior. Quanto ao estatuto profissional, 60.4% estavam empregada, 14.3% desempregadas, 17.2% eram estudantes e 4.9% eram reformados ou tinham outro estatuto profissional. O rendimento financeiro médio era de 659.84 euros por ano ($DP = 832.52$ euros).

A amostra *online* foi constituída por 278 participantes. A idade média foi de 39.61 anos ($DP = 6.19$ anos). Quanto ao estado civil, 77.3% eram casados ou união de facto, 13.3% eram solteiros, 9.0% eram divorciados ou separados e 0.4% eram viúvos. Aproximadamente, 8.3% dos participantes relataram terem uma escolaridade igual ou inferior ao nono ano, 24.1% detinham ensino secundário e 67.3% tinham um grau obtido no ensino superior. Relativamente ao estatuto profissional, 78.4% estavam empregados, 13.7% desempregados e 4% reformados. O rendimento financeiro médio era de 1167.70 euros por ano ($DP = 797.52$ euros).

A amostra papel-e-lápis foi constituída por 192 participantes. A idade média foi de 29.85 anos ($DP = 10.97$ anos). O estado civil médio era de 50.0% casados ou união de facto, 46.4% solteiro, 1.6% divorciado ou separado e 0.5% viúvo. Aproximadamente, 20.8% dos participantes relataram terem uma escolaridade igual ou inferior ao nono ano, 40.4% detinham ensino secundário e 35.9% tinham um grau obtido no ensino superior. Quanto ao estatuto profissional, 34.4% estavam empregados, 15.1% desempregados, 42.2% eram estudantes e 2.1% reformados ou tinha outro estatuto profissional. O rendimento financeiro médio era de 2.10 euros por ano ($DP = 1.23\%$ euros).

Medidas

O **Patient Health Questionnaire** (PHQ-9) é um instrumento que avalia a presença e a gravidade de sintomas depressivos (Kroenke, Spitzer, & Williams, 2001). Este é uma versão reduzida do PHQ completo. É constituído por 9 itens e refere-se às duas últimas semanas. Sendo que se deverá assinalar o grau com que foi afetado nas últimas duas semanas. Este deve utilizar a escala de *Likert* de quatro pontos (0- “*Nunca*” até 3- “*Em quase todos os dias*”). A cotação do PHQ9 vai dar uma pontuação total que pode

variar entre zero e 27 valores, sendo que a frequência dos sintomas depressivos será maior quanto maior for a pontuação.

Procedimento

O presente estudo baseou-se em dois métodos de recolha de dados (*online* e papel-lápis). Inicialmente, foi pedido a colaboração de alguns participantes para responderem ao questionário com o objetivo de estimar o tempo necessário para responder. Os participantes eram estudantes universitários em contexto sala de aula. Foram convidados a participar neste estudo, estudantes de diferentes cursos de uma universidade portuguesa. Posteriormente, foi explicado os objetivos de estudo e procedimentos de investigação, tendo sido a participação dos estudantes de forma voluntária, sem qualquer recompensa por participar. Os participantes que aceitaram colaboram neste estudo, foi entregue um consentimento informado, bem como o protocolo de investigação que foi preenchido por cada estudante.

Posteriormente, foi distribuído o questionário aos participantes, sendo pedido que respondesse individualmente, o questionário demorou aproximadamente 15 minutos a ser preenchido. Quando surgia dúvidas, estas eram tiradas individualmente pelo investigador, uma vez que este manteve-se presente para esse efeito.

No que concerne à recolha de dados *online*, foi apresentado um consentimento informado, onde explicava o objetivo de estudo, o procedimento, a confidencialidade, o tratamento de dados, bem como os contactos dos investigadores para qualquer esclarecimento que surgisse. A participação foi de forma voluntária, sem ter qualquer recompensa por participar. Foi garantido total anonimato nas respostas dadas, e os participantes deram autorização quando avançaram para a página seguinte. Todas as respostas foram armazenadas numa base de dados segura.

Cada participante teve um código que estava associado a cada consentimento informado. Só os elementos responsáveis pelo estudo, é que tinham acesso à lista que associava cada participante ao respetivo consentimento.

Os participantes, foram selecionados via *online*, lista de correio eletrónico, e divulgação online (e.g., redes sociais e meios de comunicação social). O questionário tinha um tempo estimado de 15 minutos a preencher.

Em seguida, foram efetuados os procedimentos de limpeza de dados de acordo com as recomendações Funk e Rogge (2007), que são utilizados em vários estudos efetuados *online*. Mais especificamente, os participantes que não foram considerados válidos, por não terem respondido com atenção ou dedicação e/ou, que não completaram pelo menos 70% do total do inquérito, foram eliminados.

Na base de dados, que foi utilizada pela equipa de investigação para a análise estatística, não foi mencionado qualquer informação, que possa identificar o participante. Sendo, impossível a qualquer elemento da equipa de investigação associar as respostas a um participante individualmente.

Análise estatística

Para testar a sensibilidade dos resultados comparar a estrutura fatorial do PHQ-9 com dados recolhidos *online* versus dados recolhidos papel-lápis, foi conduzida uma análise de frequência de cada um dos itens para cada uma das amostras. Para examinar a validade de construto do PHQ-9 nas duas amostras pretende-se testar diferenças sociodemográficas (idade, género, e nível de educação) entre participantes cujos dados foram recolhidos *online* e os participantes com dados recolhidos papel-lápis, para o PHQ-9, foram conduzidas, de forma individual, duas análises fatoriais exploratórias (AFE), a fim de testar a estrutura fatorial do instrumento. O método de extração utilizado foi o de componentes principais, seguida de rotação *Direct Oblim*. Considerando os princípios conceptuais da EAF, este tipo de rotação fatorial é o mais adequado, uma vez que é concetualmente esperada alguma interdependência dos sintomas de depressão (Field, 2009). A seleção de componentes principais foi conduzida através da aplicação do critério de Kaiser (1958), em que foram considerados componentes principais com eigenvalue (valor-próprio) superior a 1. Adicionalmente, o critério utilizado para saturações significativas dos itens foi de $>.30$ (Straub, 1989). O teste de esfericidade de Bartlett e a medida de Kaiser-Meyer-Olkin (KMO) foram utilizados para aferição da adequação da amostra para condução da análise factorial.

Para o objetivo três, auspiciasse testar diferenças ao nível dos *cours* totais do PHQ-9 em função do tipo de recolha de dados (*online* versus papel-lápis), foi conduzida uma análise de consistência de alfa de *Conbrach*. Utilizaram-se as recomendações de Nunnally e Bernstein (1994) para análise do alfa de *Cronbach* (e.g., Nunnally & Bernstein, 1994), em que valores iguais ou superiores a .80 indicam excelentes níveis de fiabilidade.

Finalmente, foram realizados testes *t* individuais para examinar diferenças entre as duas amostras nas médias de cada um dos itens e no *score* total do PHQ-9.

Resultados

Sensibilidade dos resultados

Com a finalidade de examinar a sensibilidade dos resultados, foi efetuada a análise de frequência para todos os itens com o objetivo de averiguar se todas as categorias de resposta estavam representadas quer na amostra que respondeu eletronicamente ao PHQ9 (Tabela 1), quer na amostra que respondeu em papel e lápis (Tabela 2).

Tabela 1. Análise das frequências dos itens do PHQ9 respondido online

Item	Escala Likert de Resposta			
	0 'Nunca'	1 'Em vários dias'	2 'Mais de metade dos dias'	3 'Quase todos os dias'
1. Tive pouco interesse ou prazer em fazer coisas	75	160	30	13
2. Senti desânimo, em baixo ou falta de esperança	109	124	31	14
3. Tive dificuldade em adormecer ou em dormir sem interrupções, ou dormi demais	89	117	41	31
4. Senti cansaço ou falta de energia	35	149	47	37
5. Tive falta ou excesso de apetite	111	104	31	30
6. Senti que não gosto de mim próprio/a — ou que sou um(a) falhado/a ou me desiludi a mim próprio/a ou à minha família	185	61	18	13
7. Tive dificuldade em concentrar-me nas coisas, como ao ler o jornal ou ver televisão	142	106	18	9
8. Movimentei-me ou falei tão lentamente que outras pessoas poderão ter notado. Ou o oposto: estive agitado/a a ponto de andar de um lado para o outro muito mais do que é habitual	198	56	16	4
9. Pensei que seria melhor estar morto/a, ou em magoar-me a mim próprio/a de alguma forma	256	11	9	2

Tabela 2. *Análise das frequências dos itens do PHQ9 respondido papel-e-lápis*

Item	Escala Likert de Resposta			
	0 'Nunca'	1 'Em vários dias'	2 'Mais de metade dos dias'	3 'Quase todos os dias'
1. Tive pouco interesse ou prazer em fazer coisas	62	100	17	10
2. Senti desânimo, em baixo ou falta de esperança	76	87	19	5
3. Tive dificuldade em adormecer ou em dormir sem interrupções, ou dormi demais	57	75	32	24
4. Senti cansaço ou falta de energia	31	99	41	16
5. Tive falta ou excesso de apetite	91	59	23	16
6. Senti que não gosto de mim próprio/a — ou que sou um(a) falhado/a ou me desiludi a mim próprio/a ou à minha família	139	35	10	5
7. Tive dificuldade em concentrar-me nas coisas, como ao ler o jornal ou ver televisão	101	61	20	5
8. Movimentei-me ou falei tão lentamente que outras pessoas poderão ter notado. Ou o oposto: estive agitado/a a ponto de andar de um lado para o outro muito mais do que é habitual	123	51	13	2
9. Pensei que seria melhor estar morto/a, ou em magoar-me a mim próprio/a de alguma forma	172	13	2	2

Para ambas as amostras, verificou-se que todas as categorias de respostas foram representadas em todos os itens, sugerindo que os todos os itens têm sensibilidade para distinguir os participantes, quer na amostra *online*, quer na amostra papel-e-lápis. A análise de assimetria e curtose do total do PHQ9 revelaram que este se situava entre os valores de -

1 e +1, atestando a normalidade das distribuições, quer na amostra *online*, quer na amostra papel-e-lápis.

Validade de construto

Amostra *online*. Os testes de esfericidade de Bartlett e KMO revelaram que esta amostra era estatisticamente adequada para condução da AFE. A AFE revelou que apenas um componente apresentava um valor eigenvalue superior a 1 (4.7), sugerindo a extração de um único fator. Este fator único explicava 52.4% da variância total. As saturações dos itens foram elevadas, variando entre .54 e .83 (Tabela 3).

Tabela 3. *Coefficientes de saturação standardizados dos itens da solução fatorial do PHQ9 para amostra online*

Item	Saturação
1. Tive pouco interesse ou prazer em fazer coisas	.77
2. Senti desânimo, em baixo ou falta de esperança	.83
3. Tive dificuldade em adormecer ou em dormir sem interrupções, ou dormi demais	.71
4. Senti cansaço ou falta de energia	.80
5. Tive falta ou excesso de apetite	.68
6. Senti que não gosto de mim próprio/a — ou que sou um(a) falhado/a ou me desiludi a mim próprio/a ou à minha família	.79
7. Tive dificuldade em concentrar-me nas coisas, como ao ler o jornal ou ver televisão	.72
8. Movimentei-me ou falei tão lentamente que outras pessoas poderão ter notado. Ou o oposto: estive agitado/a a ponto de andar de um lado para o outro muito mais do que é habitual	.64
9. Pensei que seria melhor estar morto/a, ou em magoar-me a mim próprio/a de alguma forma	.54

Amostra papel-e-lápis. Os testes de esfericidade de Bartlett e KMO revelaram que esta amostra mostrava-se estatisticamente adequada para a realização da AFE. Os resultados da EFA demonstraram a existência de único componente com um valor eigenvalue superior a 1 (4.4), o que sugeriu a adopção de uma estrutura unifatorial. Este fator único explicava 49.8% da variância total. Tal como pode ser verificado na Tabela 4, as saturações dos itens foram elevadas, variando entre .54 e .83.

Tabela 4. *Coefficientes de saturação standardizados dos itens da solução fatorial do PHQ9 para amostra papel-e-lápis*

Item	Saturação
1. Tive pouco interesse ou prazer em fazer coisas	.67
2. Senti desânimo, em baixo ou falta de esperança	.79

3. Tive dificuldade em adormecer ou em dormir sem interrupções, ou dormi demais	.69
4. Senti cansaço ou falta de energia	.75
5. Tive falta ou excesso de apetite	.67
6. Senti que não gosto de mim próprio/a — ou que sou um(a) falhado/a ou me desiludi a mim próprio/a ou à minha família	.74
7. Tive dificuldade em concentrar-me nas coisas, como ao ler o jornal ou ver televisão	.70
8. Movimentei-me ou falei tão lentamente que outras pessoas poderão ter notado. Ou o oposto: estive agitado/a a ponto de andar de um lado para o outro muito mais do que é habitual	.70
9. Pensei que seria melhor estar morto/a, ou em magoar-me a mim próprio/a de alguma forma	.65

Fiabilidade

Para avaliar a fiabilidade do PH9 em ambas as amostras, examinou-se a consistência interna do instrumento, através da análise dos alfas de *Cronbach*. Para a amostra que respondeu ao PHQ9 de forma eletrónica, o alfa de *Cronbach* obtido foi de .88. Para a amostra papel-e-lápis, o alfa de *Cronbach* situou-se em .86. Tendo em consideração as recomendações para análise dos alfa de *Cronbach* (e.g., Nunnally & Bernstein, 1994), o PHQ9 apresentou excelentes níveis de fiabilidade em ambas as amostras.

Diferenças nas médias

Foram testadas diferenças entre as duas amostras nas médias obtidas em cada um dos nove itens, assim como no *score* total do PHQ9. As médias por amostra e os resultados dos testes *t* estão descritos na Tabela 5. Tal como se pode verificar, não foram encontradas diferenças estatisticamente em cada um dos itens e no *score* total do PHQ9 entre a amostra *online* e amostra papel-e-lápis.

Tabela 5. *Diferenças entre as amostras nos itens e total do PHQ9*

	Amostra <i>Online</i> (<i>n</i> = 278)	Amostra papel-e-lápis (<i>n</i> = 192)	<i>t</i>
	<i>M</i> (<i>DP</i>)	<i>M</i> (<i>DP</i>)	
1. Tive pouco interesse ou prazer em fazer coisas	0.93 (0.75)	0.87 (0.78)	.89
2. Senti desânimo, em baixo ou falta de esperança	0.82 (0.82)	0.75 (0.75)	.96
3. Tive dificuldade em adormecer ou em dormir sem interrupções, ou dormi demais	1.05 (0.95)	1.12 (0.99)	-.79
4. Senti cansaço ou falta de energia	1.31 (0.86)	1.22 (0.83)	1.06

5. Tive falta ou excesso de apetite	0.93 (0.97)	0.81 (0.95)	1.29
6. Senti que não gosto de mim próprio/a — ou que sou um(a) falhado/a ou me desiludi a mim próprio/a ou à minha família	0.49 (0.82)	0.37 (0.70)	1.65
7. Tive dificuldade em concentrar-me nas coisas, como ao ler o jornal ou ver televisão	0.61 (0.75)	0.62 (0.78)	-.08
8. Movimentei-me ou falei tão lentamente que outras pessoas poderão ter notado. Ou o oposto: estive agitado/a a ponto de andar de um lado para o outro muito mais do que é habitual	0.36 (0.66)	0.44 (0.67)	-1.18
9. Pensei que seria melhor estar morto/a, ou em magoar-me a mim próprio/a de alguma forma	0.12 (0.13)	(0.44) (0.46)	.10
<i>Score total</i>	6.62 (5.13)	6.26 (4.76)	.76

Discussão

O presente estudo tinha como objetivo principal avaliar as respostas *online* e de papel-e-lápis numa medida para compreender se as qualidades psicométricas do PHQ-9 sofriam alterações.

Os resultados deste estudo revelaram que nas duas amostras que foram testadas não foram encontradas diferenças estatisticamente, nem no *score* total do PHQ-9. Quanto à validade do construto para amostra em papel, revelou que este era estatisticamente adequado para a condução da realização da AFE. O mesmo se verificou na validade do construto para amostra *online*. Relativamente à fiabilidade do PHQ-9 em ambas amostras e após examinar a consistência interna do instrumento, verificou-se que o PHQ-9 tinha uma excelente fiabilidade em ambas amostras.

Este estudo tinha quatro objetivos e cada objetivo tinha a sua hipótese de trabalho, sendo que foram as quatro hipóteses confirmados no presente estudo.

A sensibilidade dos resultados seria demonstrada em ambas as amostras, sendo que no presente estudo todas as categorias foram representadas, confirmando assim a primeira hipótese de trabalho, que pretendia que a sensibilidade de respostas fossem confirmadas em todas as categorias de resposta do PHQ-9. Relativamente à validade de construto, esta não apresenta diferenças significativas, tendo sido confirmada a hipótese de estudo. Segundo Donker (2010), num estudo sobre três medidas de avaliação da depressão através da internet, revelam através dos resultados obtidos que não existem valores significativos a nível da validade de construto.

A versão *online* e a versão papel-lápis não sofreram alterações entre as respostas dos participantes de ambas as versões, uma vez que os valores de Alfa de *Cronbach* foram similares. No estudo conduzido por Carlbring et al. (2007) confirmam os mesmos resultados que o presente estudo relativamente ao Alfa de *Cronbach*, não sofrendo alterações significativas entre as duas versões.

A recolha de dados seria um aspeto importante e a ter em conta neste estudo. Para tal, pretendia-se que não houvesse diferenças nos *scores*. Posteriormente, foi confirmado que no presente estudo não foram encontradas diferenças estatisticamente nos *scores* do PHQ-9. O Hollandare, Askerlund, Nieminen & Engstrom (2008) conduziram um estudo para verificar se o instrumento de avaliação BDI-II sofriam alterações psicométricas, e os resultados obtidos levam a concluir que este instrumento poderia ser utilizado na versão *online*, uma vez que não sofrem alterações psicométricas, indicando que o mesmo não revela diferenças significativas entre as duas versões. Assim como o BDI-II, no estudo de Hollandare et al., (2008), o PHQ-9 também pode ser utilizado na versão *online*, uma vez que não sofre alteração psicométrica. No estudo de Zlomke, (2009) foram investigadas as propriedades psicométricas de duas medidas, o *Penn State Worry Questionnaire* e a Escala de Depressão, Ansiedade e Stress, para a versão *online*. Os resultados revelam que as duas medidas podem ser usadas na versão *online* em termos de confiabilidade e validade. O estudo conduzido por Spek et al., (2008) conclui que as propriedades psicométricas do EDS quando administradas na versão *online* são comparáveis à versão papel-lápis, ou seja, não sofrem alterações. Estes resultados vão de encontro a estudos anteriores (Vallejo, Mañanes, Comeche & Díaz, 2008; Hollandare et al., 2008) em que não foram encontradas diferenças estatisticamente significativas entre as duas versões.

No que concerne à prática clínica, esta forma de avaliação teria as suas desvantagens, uma vez que a avaliação *online* iria implicar que os profissionais não se deslocassem para locais apropriados, como clínicas, pois poderiam realizar o seu trabalho a partir de casa. Teria um contributo negativo para a economia, pois iria tirar muitos locais de trabalho. Por outro, quem estivesse interessado neste método, teria de ter algumas condições económicas para poder usufruir. Para recorrer a este método, teriam de ter um computador que tivesse um bom processamento para que não fosse aborrecido e demorasse muito tempo a responder aos questionários e teriam de ter obrigatoriamente internet (Buchanan, 2003; Sheehan e McMillan, 1999).

Neste estudo, existem três limitações importantes. —Em primeiro, teria sido relevante ter emparelhado as duas amostras nas principais características demográficas.

Apesar de não ter havido diferenças no comportamento psicométrico da escala nas duas versões, em caso de terem ocorrido, elas poderiam ter sido devido ao não total controlo das características demográficas entre as duas amostras. Em segundo, o estudo foi conduzido exclusivamente com participantes mulheres. Apesar de criar homogeneidade nos resultados, que contribui para o poder estatístico das análises, os resultados presentes não são generalizáveis a homens. Estudos posteriores devem incluir na amostra participantes do sexo masculino e testar diferenças em função do género. Finalmente, não foram incluídas no protocolo medidas adicionais que permitissem a avaliação da validade convergente e divergente do PHQ-9 nas suas versões *online* e papel-lápis. Estudos posteriores devem suprir esta limitação, a fim de dar mais robustez metodológica aos resultados do presente estudo.

Por outro lado, a maior força deste estudo foi testar pela primeira vez em Portugal o comportamento psicométrico de uma medida nas duas versões *online* e papel-lápis.

Em Portugal são escassos os estudos desenvolvidos no âmbito da avaliação *online*, tornando-se uma força para a realização deste estudo, pois irá contribuir para que haja mais estudos e mais informação relativamente a este tema. Contudo, seria necessário e importante apostar mais em estudos relacionados com avaliação *online*, para aprofundar mais e ter resultados mais significativos relativamente a outras medidas psicométricas. Seria interessante perceber como as pessoas iriam reagir ao método da avaliação psicológica ser realizada na versão *online*, perceber se iriam sentir à vontade com o facto de não estar presente fisicamente um técnico especializado da área, mas sim através de um computador.

No futuro, possíveis investigações poderiam estar relacionadas com o número de participantes serem igual nas duas versões, bem como o método de recolha ser equivalente na amostra e não diferir no estatuto profissional.

Referências

- Ballegooijen, W. B., Riper, H. R., Cuijpers, P. C., Oppen, P. O., Smit, J. S. (2016). Validation of online psychometric instruments for common mental health disorders: a systematic review. *BMC Psychiatry*, 16-45.
- Bjornsdotter, A., Enebrink, P., Ghaderi, A. (2013). Psychometric properties of online administered parental strengths and difficulties questionnaire (SDQ), and normative

- data based on combined online and paper-and-pencil administration. *Child and Adolescent Psychiatry and Mental Health*, <https://doi.org/10.1186/1753-2000-7-40>.
- Buchanan, T. (2003). Internet-based Questionnaire Assessment: Appropriate Use in Clinical Contexts. *Cognitive Behaviour Therapy*, 32, , 100-109.
- Burke, J., Burke, K., & Baker, J. . (1995). Test retest reliability in psychiatric patients of the SF-36 Health Survey. *International Journal of Methods in Psychiatric Research*, 189-194.
- Coyne. (1999). The impact of mode of data collection on data quality with subjects having low literacy skills (African American women). *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 1552.
- Campos, J.C., Zucoloto, M.Z., Bonafé, F. B., Jordani, P. C., Maroco, J. M. (2011). Reliability and validity of self-reported burnout in college students: A cross randomized comparison of paper-and-pencil vs. online administration. *Computers in Human Behavior*, Vol.27, 1875-1883.
- Carlbring, P. C., Brunt, S. B., Bohman, S. B., Austin, D. A., Richards, J. R., Goran, L. G., Andersson, G. A. (2007). Internet vs. Paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, Vol. 23, 1421-1434.
- Donker, T., Straten, V.A., Marks, I., Cuijpers, P.(2010). Brief self-rated screening for depression on the Internet. *Journal of Affective Disorders*, Vol.122,, 253-259.
- Fouladi, T.R., McCarthy, C.J., Moller, N. (2002). Paper-and-Pencil Or Online? Evaluating mode effects on measures of emotional functioning and attachment. *Sage journals*, Vol. 9, 204-215 .
- Hollandare, F., Askerlund, M.A., Nieminen, A., Engstrom, I. (2008). Can the BDI-II and MADRS-S be Transferred to Online Use Without Affecting their Psychometric Properties?. *E-Journal of Applied Psychology*:4, 2,, 63-65.
- Luce, K. H., Winzelberg, A. J., Das, S., Osborne, M. I., Bryson, S. W., & Taylor, C. B . (2007). Reliability of self-report: Paper versus online administration. . *Computers in Human Behavior*, 1384-1389.
- Kessler, R., C. & Bromet, E. J . (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, Vol. 34, 119-138
- Kroenke, K., Spitzer, R. L., & Williams, J. B., (2007). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613.

- Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., et al. (2002). Test-retest reliability of alcohol measures: Is there a difference between internet-based assessment and traditional methods? *Psychology of Behaviors*, 56-61.
- Naus, M.N., Philipp, L.P., Samsi, M. S. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior*, 25,, 1-7.
- Nguyen, DP., Klein, B., Meyer, D., Austin, DW., Abbott, JA. (2015). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *J Med Internet Res*, 21;17 (9), 218. Doi:10.2196/jmir.4195.
- Reips, U. D. (2001). The web experimental psychology lab: Five years of data collection on the Internet. *Behavior Research Methods Instruments & Computers*, 33,, 201-211.
- Riva, G., Teruzzi, T., & Anolli, L. (2004). The use of the internet in psychological research: Comparison of online and offline questionnaires. *Cyberpsychology & Behavior*, 6 (1)73-80.
- Luce, K. H., Winzelberg, A. J., Das, S., Osborne, M. I., Bryson, S. W., & Taylor, C. B . (2007). Reliability of self-report: Paper versus online administration. . *Computers in Human Behavior*, 1384-1389.
- Kessler, R., C. & Bromet, E. J . (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, Vol. 34, 119-138
- Kroenke, K., Spitzer, R. L., & Williams, J. B., (2007). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606-613.
- Miller, E. T., Neal, D. J., Roberts, L. J., Baer, J. S., Cressler, S. O., Metrik, J., et al. (2002). Test-retest reliability of alcohol measures: Is there a difference between internet-based assessment and traditional methods? *Psychology of Behaviors*, 56-61.
- Naus, M.N., Philipp, L.P., Samsi, M. S. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior*, 25,, 1-7.
- Nguyen, DP., Klein, B., Meyer, D., Austin, DW., Abbott, JA. (2015). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *J Med Internet Res*, 21;17 (9), 218. Doi:10.2196/jmir.4195.

- Reips, U. D. (2001). The web experimental psychology lab: Five years of data collection on the Internet. *Behavior Research Methods Instruments & Computers*, 33,, 201-211.
- Riva, G., Teruzzi, T., & Anolli, L. (2004). The use of the internet in psychological research: Comparison of online and offline questionnaires. *Cyberpsychology & Behavior*, 6 (1)73-80.
- Ryan, J. M., Corry, J. R., & Attewell, R. . (2002). A comparison of an electronic version of the SF-36 General Health Questionnaire to the standard paper version. *Quality of Life Research*, 11, , 19-26.
- Sheehan, K. B., McMillan, S. J. (1999). Response variation in e-mail surveys: Na exploration. *Journal of Advertising Research*, 45-54.
- Vallejo, M. V., Jordán, C.J., Díaz, M. D., Comeche, M. C., Ortega, J. O. (2007). Psychological assessment via the internet: A reliability and validity study of online (vs paper-and- pencil) versions of the general health questionnaire-28 (GHQ-28) and the symptoms check-list-90-revised (SCL-90-R). *Journal of Medical Internet Researche*, 9(1), , 1-2.
- Vallejo, M. V., Mañanes, G. M., Comeche, M. I. C., Díaz, M. D. (2008). Comparison between administration via internet and paper-and- pencil administration of two clinical instruments: SCL-90-R and GHQ-28. *Journal of Behavior Therapy*, 39, , 201-208.
- Ward, P., Clark, T., Zabriskle, R., Morris, T. (2014). Paper/Pencil Versus Online Data Collection. *Journal of Leisure Research*, 46 (1), 84-105.
- Wood, E., Nosko, A., Desmarais, S., Ross, C., Irvine, C. (2006). Online and traditional paper-and-pencil survey administration:Examining experimental presence, sensitive material and long surveys. *The Canadian Journal of Human sexuality*, 15 (3-4), 147-155.