

MỐI QUAN HỆ GIỮA ĐỘ KHÓ VÀ ĐỘ PHÂN CÁCH CỦA CÂU HỎI TRẮC NGHIỆM SINH LÝ HỌC TRONG CÁC ĐỀ THI TUYỂN SINH SAU ĐẠI HỌC DÀNH CHO ĐỐI TƯỢNG CHUYÊN KHOA CẤP I TỪ NĂM 2018 ĐẾN NĂM 2022 TẠI ĐẠI HỌC Y DƯỢC THÀNH PHỐ HỒ CHÍ MINH

Phạm Dương Uyển Bình¹, Trần Thị Diệu¹, Nguyễn Đình Chương¹, Mai Phương Thảo¹, Nguyễn Anh Vũ¹, Vĩnh Sơn¹, Phạm Lê An¹

TÓM TẮT

Mục tiêu: Đánh giá chính xác năng lực của thí sinh là điều quan trọng để đánh giá kiến thức và kỹ năng của họ trong các lĩnh vực cụ thể. Giả định truyền thống cho rằng các câu hỏi dễ trong bài kiểm tra sẽ có độ phân cách thấp, hạn chế trong việc phân biệt giữa học viên ở các cấp độ thành thạo khác nhau. Tuy nhiên, có những bằng chứng cho thấy khả năng phân cách của câu hỏi có thể bị ảnh hưởng bởi các yếu tố khác ngoài độ khó của câu như đặc điểm của mẫu. Do đó, nghiên cứu này phân tích câu hỏi trắc nghiệm của bài kiểm tra tuyển sinh sau đại học về độ khó và phân cách nhằm mục đích phân tích mối liên hệ phức tạp của độ khó và độ phân cách trong ngữ cảnh của bài thi môn Sinh lý học tuyển vào chương trình đào tạo bác sĩ Chuyên khoa cấp I.

Phương pháp: Nghiên cứu hồi cứu dọc được áp dụng để phân tích toàn diện bài kiểm tra Sinh lý học gồm 120 câu, được thực hiện trong vòng 5 năm, từ năm 2018 đến 2022. Nghiên cứu đã thu thập một tập dữ liệu bao gồm 2268 kết quả kiểm tra từ một nhóm đa dạng thí sinh dự thi sau đại học. Phân tích câu đã được thực hiện để xác định độ khó và phân cách trên phần mềm phân tích Basicstat. Ngoài ra, các phép kiểm dùng để so sánh số câu trả lời đúng giữa các đặc điểm mẫu khác nhau, SPSS 20 để xếp loại tốt nghiệp và năm tốt nghiệp.

Kết quả: Phần lớn trong số 120 câu của bài thi môn Sinh lý học được phân loại là dễ hoặc rất dễ. Trong số các câu này, 30,8% đến 43,3% câu hỏi trong vòng 5 năm có chỉ số phân cách r_{pbis} trên 0,3. Kết quả phân tích đặc điểm của mẫu và số câu làm đúng cho thấy sự khác biệt có ý nghĩa về số câu trả lời đúng giữa các thí sinh tốt nghiệp từ các trường đại học khác nhau [$F(5,3921) = 36,694, p < 0,001$], xếp loại tốt nghiệp [$F(5,3921) = 241,237, p < 0,001$] và năm tốt nghiệp khác nhau [$F(3,3923) = 36,082, p < 0,001$].

Kết luận: Nghiên cứu chứng minh rằng các câu hỏi dễ trong bài kiểm tra không nhất thiết có độ phân cách thấp. Điều này làm nổi bật mối quan hệ phức tạp giữa độ khó và độ phân cách câu. Sự chênh lệch đáng kể về năng lực của mẫu thí sinh có thể

¹ Đại học Y Dược Thành phố Hồ Chí Minh

Thông tin liên lạc: **TS. Phạm Dương Uyển Bình**, ĐT: 0961156505 Email: uyenbinh@ump.edu.vn

gia tăng độ phân cách của câu. Do đó, ban ra đề thi cần nhận thức rằng chỉ dựa vào độ khó của câu hỏi thi chưa thể kết luận khả năng phân loại năng lực của thí sinh, người học.

Từ khóa: độ khó, logit, độ phân cách, r_{pbis} , CTT, IRT, Rasch.

ABSTRACT

EXPLORING THE COMPLEX RELATIONSHIP BETWEEN ITEM DIFFICULTY AND DISCRIMINATION INDEX IN PHYSIOLOGY MCQS IN THE POST-GRADUATE ADMISSION TEST FOR 1ST SPECIALIST FROM 2018 TO 2022 AT UNIVERSITY OF MEDICINE AND PHARMACY IN HO CHI MINH

Pham Duong Uyen Binh, Tran Thi Dieu, Nguyen Dinh Chuong, Mai Phuong Thao, Nguyen Anh Vu, Vinh Son, Pham Le An

Background: Assessing students' abilities accurately is essential for evaluating their knowledge and skills in specific disciplines. The conventional assumption suggests that easy items on tests exhibit low discrimination, implying limited ability to differentiate among students of varying proficiency levels. However, there are evidences that item discriminative power can be affected by factors other than item difficulty such as sample characteristics. Therefore, this study aims to explore the intricate relationship between difficulty and discrimination indices in the context of Physiology post-graduate admission test through MCQ item analysis.

Method: A longitudinal study was employed to thoroughly analyze a 120-item Physiology test administered over a span of 5 years, from 2018 to 2022. The study gathered a comprehensive dataset of 2268 test responses from a diverse group of post-graduate students. Retrospective item analysis was conducted to carefully examine item difficulty and discrimination indices using BasicStat software. Additionally, One-way ANOVA was used to compare the numbers of correct answers among different sample characteristics, such as graduation ranks, or years of graduation using SPSS 20.

Results: Most of the 120 items on the Physiology test were classified as easy or very easy. Among these easy items, a significant proportion, ranging from 30,8% to 43,3% over 5 years, had a r_{pbis} discrimination index above 0,3. The results of the analysis of sample characteristics and the number of correct answers showed significant differences in the number of correct responses between students from different schools [$F(5,3921) = 36,694, p < 0,001$], graduation ranks [$F(5, 3921) = 241,237, p < 0,001$] and graduation years [$F(3,3923) = 36,082, p < 0,001$].

Conclusion: Our study demonstrates that easy items in the Physiology test do not necessarily indicate low discrimination of students' ability. It highlights the intricate relationship between item difficulty and discrimination power. The significant variation in the abilities of the sample candidates may increase the

discrimination power of the questions. Therefore, educators and test developers should recognize that item difficulty alone does not determine discrimination potential.

Keywords: item difficulty, item 27% discrimination, r_{pbis} , reliability, Classical Test Theory, Items Response Theory, Rasch.

ĐẶT VẤN ĐỀ

Đánh giá năng lực người học là một khía cạnh cơ bản của thực hành giáo dục, cung cấp những hiểu biết có giá trị về kiến thức và kỹ năng của người học. Các bài kiểm tra thành tích nói riêng đóng vai trò quan trọng trong việc đo lường kết quả học tập của người học và hiểu mức độ thành thạo của họ trong các môn học khác nhau. Các khía cạnh quan trọng khi thiết kế các bài kiểm tra là độ khó và độ phân cách của các câu hỏi, vì chúng ảnh hưởng trực tiếp đến giá trị và độ tin cậy của bài đánh giá.

Theo quan niệm truyền thống, các câu hỏi dễ trong bài kiểm tra là dấu hiệu của độ phân cách thấp, không phân biệt được các thí sinh có trình độ khác nhau [1]. Ngược lại, các câu hỏi khó được cho là có khả năng phân biệt cao hơn giữa nhóm người học có thành tích tốt so với nhóm còn lại [1]. Tuy nhiên, các nghiên cứu gần đây bắt đầu chứng minh rằng mối quan hệ giữa độ khó của câu hỏi và chỉ số phân biệt phức tạp hơn [2]. Nghiên cứu này nhằm khám phá mối quan hệ giữa độ khó của câu hỏi và chỉ số phân biệt thông qua bài kiểm tra môn Sinh lý học.

Môn Sinh lý học thuộc khối kiến thức y học cơ sở, có vai trò quan trọng đối với người học thuộc khối ngành khoa học sức khỏe, giải thích hoạt động chức năng của cơ thể ở trạng thái bình thường. Môn Sinh lý học được giảng dạy trong giai đoạn tiền lâm sàng; và được vận dụng để giải thích cơ chế bệnh lý, triệu chứng bệnh, làm tiền đề cho việc chẩn đoán và điều trị trong đào tạo những năm học lâm sàng về sau. Vì vậy, bài thi môn Sinh lý học được lựa chọn để làm môn thi tuyển sinh đánh giá năng lực đầu vào của người dự tuyển vào các bậc đào tạo sau đại học.

Các tham số câu hỏi bao gồm độ khó và độ phân cách. Độ khó của câu hỏi trong lý thuyết đáp ứng câu hỏi (Items Response Theory) là một tham số thể hiện mức độ khó của một câu hỏi đối với người làm bài, thường được tính bằng điểm trên thang đo năng lực mà người làm bài có 50% khả năng trả lời đúng câu hỏi. Các mô hình IRT giả định rằng khả năng thí sinh (TS) trả lời đúng câu hỏi tăng lên khi năng lực của họ tăng lên. *Vi dụ, một câu hỏi có độ khó cao là một câu hỏi mà chỉ TS có năng lực cao mới có khả năng trả lời đúng.* **Độ khó** của câu hỏi là một tham số quan trọng trong IRT vì nó cho phép tạo ra các bài kiểm tra được điều chỉnh theo mức độ năng lực của người làm bài. Lý tưởng nhất là, nếu hệ thống phân cấp câu hỏi đại diện cho một phạm vi năng lực đầy đủ và các câu hỏi thường sẽ dao động từ -2 đến $+2$ logit với giá trị logit càng thấp thì câu hỏi càng dễ [3]. Bên cạnh đó, chỉ số phân cách (r_{pbis}) trên 0,3 cho thấy sự phân biệt

hiệu quả giữa các ứng viên mạnh và yếu hơn, góp phần vào khả năng đánh giá năng lực tổng thể của thí sinh thông qua kỳ thi.

Theo truyền thống, người ta cho rằng các câu hỏi dễ có độ phân cách kém để phân biệt giữa những người học có mức độ hiểu biết khác nhau. Tuy nhiên, các nghiên cứu gần đây cho thấy rằng giả định này có thể không đúng đối với tất cả các trường hợp. Karadag et al., 2016 cho thấy các câu hỏi ở mức độ nhớ và hiểu mặc dù được đánh giá “dễ” hơn so với câu hỏi ở mức độ áp dụng, nhưng lại có thể phân biệt tốt hơn giữa nhóm TS điểm cao và thấp [3]. Ngoài ra còn có thể có những yếu tố cơ bản khác góp phần vào khả năng phân biệt năng lực người học. Do đó, nghiên cứu của chúng tôi nhằm khám phá mối quan hệ giữa độ khó và phân cách trong bộ đề thi trắc nghiệm của môn thi Sinh lý học. Việc hiểu rõ mối tương quan của độ khó và độ phân biệt của câu hỏi giúp đề xuất các chiến lược đánh giá năng lực của người học.

ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Thiết kế nghiên cứu

Một nghiên cứu hồi cứu dọc đã được sử dụng để phân tích bài kiểm tra Sinh lý học 120 câu được thực hiện trong khoảng thời gian 5 năm, từ năm 2018 đến năm 2022. Dữ liệu thứ cấp dưới dạng phân tích câu hỏi đã được lấy từ kết quả thi. Nghiên cứu đã thu thập một tập dữ liệu toàn diện gồm 2268 bài kiểm tra từ một nhóm học viên sau đại học đa dạng, và việc phân tích tập trung vào các câu hỏi trắc nghiệm một đáp án với mỗi câu có 4 lựa chọn. Học viên sau đại học là sinh viên tốt nghiệp từ nhiều trường đại học y khoa ở Việt Nam từ trước năm 2013 đến 2022.

Công cụ

Các câu hỏi trắc nghiệm Sinh lý học được sử dụng trong kỳ thi tuyển sinh sau đại học trong 5 năm từ 2018 đến 2022 đã được phân tích. Chỉ số độ khó trong mô hình IRT (trên thang đo logit) và chỉ số phân biệt (r_{pbis}) đã được sử dụng để đánh giá các câu hỏi trắc nghiệm.

r_{pbis} : $r_{pbis} \geq 0,3$ được cho là có chỉ số phân biệt tốt.

Độ khó của các câu hỏi trắc nghiệm Sinh lý học:

Độ khó câu hỏi theo Lý thuyết cổ điển (CTT): Độ khó p bằng tỷ số phần trăm thí sinh làm đúng câu hỏi trên tổng số thí sinh tham gia làm câu hỏi đó. Độ khó p cho biết tỷ lệ thí sinh trả lời đúng câu hỏi nhiều hay ít. Giá trị p càng lớn thì câu hỏi càng dễ:

$p = \text{số thí sinh làm đúng} / \text{số thí sinh làm bài}$.

Độ khó dựa trên thang logit theo Lý thuyết ứng đáp câu hỏi (IRT): Lý tưởng nhất là, nếu hệ thống phân cấp câu hỏi đại diện cho một phạm vi năng lực đầy đủ và các câu hỏi thường sẽ dao động từ -2 đến $+2$ logit. Logit là đơn vị đo lường thang đo

khoảng dọc theo thang đo khoảng thu được từ phân tích Rasch. Dựa trên thang đo logit, chúng tôi tập trung vào 2 phân loại chỉ số độ khó bao gồm câu hỏi dễ (thang đo logit từ -0,1 đến -1); câu hỏi rất dễ (thang đo logit nhỏ hơn -1).

Thống kê

Thống kê mô tả đã được sử dụng để phân tích dữ liệu. Chỉ số độ khó theo thang đo logit và chỉ số phân biệt r_{pbis} được tính toán bằng phần mềm BasicStat.

KẾT QUẢ NGHIÊN CỨU

Qua 5 năm, đề thi tuyển sinh sau đại học môn Sinh lý học có số câu hỏi ổn định 120 câu với 4 lựa chọn. Số thí sinh dự thi tăng nhiều qua 5 năm từ 361 thí sinh năm 2018 đến 817 thí sinh năm 2022.

Bảng 1: Đặc điểm dân số nghiên cứu.

Năm	2018	2019	2020	2021	2022
Số thí sinh	361	452	678	592	817
Tổng số câu	120	120	120	120	120
Số lựa chọn	4	4	4	4	4
Xếp loại tốt nghiệp bậc đại học					
Giỏi			81		
Khá			518		
Trung bình khá			364		
Trung bình			29		
Năm tốt nghiệp đại học					
2016 về trước			258		
Năm 2017			160		
Năm 2018			311		
2019 về sau			263		

Bảng 2: Tỷ lệ câu hỏi dễ và rất dễ có độ phân cách $r_{pbis} \geq 0,3$.

Năm	$r_{pbis} \geq 0,3$ (N=120)					
	Rasch			CTT		
	-1~-0,1	<-1	Tỷ lệ (%)	0,6~0,8	$\geq 0,8$	Tỷ lệ (%)
2018	28	09	30,8	19	44	52,5
2019	31	15	38,3	26	44	58,3
2020	29	14	35,8	27	45	60
2021	36	13	40,8	24	62	71,6
2022	33	19	43,3	26	52	65

Một lượng câu hỏi dễ và rất dễ theo CTT (trong vòng 5 năm, có chỉ số phân biệt r_{pbis} tốt $\geq 0,3$), dao động từ 52,5% đến 71,6% tổng số 120 câu. Dựa trên độ khó theo thang logit, câu hỏi dễ (độ khó logit thấp từ -1 đến -0,1) và rất dễ (độ khó logit rất thấp dưới -1) dao động từ 30,8% đến 43,3% trong vòng 5 năm, có chỉ số phân biệt r_{pbis} tốt ($\geq 0,3$). Điều này cho thấy sự khác biệt trong số câu có độ phân cách tốt tính theo cách truyền thống và thang logit.

Bảng 3: Sự khác biệt về số câu đúng với các đặc điểm của thí sinh.

	M	SD	Trung bình khác biệt	Khoảng tin cậy	p value
Trường tốt nghiệp đại học					
Trường tốt nghiệp 1	91,80	15,01			
Trường tốt nghiệp 2	86,16	17,97	5,64*	3,25 ~ 8,05	<0,001
Trường tốt nghiệp 3	97,36	13,52	-5,56*	-8,08 ~ -3,04	<0,001
Trường tốt nghiệp 4	92,08	16,05	-0,27	-2,41 ~ 1,86	0,999
Trường tốt nghiệp 5	87,25	17,68	4,55*	1,90 ~ 7,20	<0,001
Trường tốt nghiệp 6	86,99	18,00	4,81*	2,38 ~ 7,25	<0,001
Xếp loại tốt nghiệp bậc đại học					
Giỏi	102,01	10,77			
Khá	95,19	13,89	0,54	0,38 ~ 0,71	<0,001
Trung bình khá	84,69	17,09	1,23	1,06 ~ 1,39	<0,001
Trung bình	73,20	17,75	1,88	1,61 ~ 2,15	<0,001
Năm tốt nghiệp đại học					
2018 - 2022	92,32	15,75			
2016 - 2017	93,22	14,88	-0,89	-2,69 ~ -0,89	0,570
2014 - 2015	88,58	17,41	3,74*	1,79 ~ 5,69	<0,001
2013 về sau	86,38	18,22	5,94*	3,90 ~ 7,98	<0,001

Phân tích phương sai một chiều (One-way ANOVA) được thực hiện để đánh giá mối quan hệ giữa xuất thân trường tốt nghiệp đại học, xếp loại tốt nghiệp, năm tốt nghiệp và số câu trả lời đúng.

Kết quả phân tích cho thấy có sự khác biệt đáng kể giữa các nhóm trường tốt nghiệp đại học [$F(5,3921) = 36,694, p < 0,001$]. Kiểm tra Tukey HSD sau kiểm định cho thấy số câu trả lời đúng của thí sinh từ trường tốt nghiệp (TN) đại học 2, 5 và 6 thấp hơn đáng kể so với trường TN 1 ($p < 0,001$).

Kết quả phân tích cũng cho thấy có sự khác biệt đáng kể giữa các nhóm loại tốt nghiệp [$F(5, 3921) = 241,237, p < 0,001$]. Kiểm tra Tukey HSD sau kiểm định cho thấy thí sinh sau đại học có loại tốt nghiệp càng cao thì có số câu trả lời đúng nhiều hơn đáng kể so với những thí sinh có loại tốt nghiệp thấp hơn ($p < 0,001$).

Ngoài ra, kết quả phân tích cũng cho thấy có sự khác biệt đáng kể giữa các nhóm năm tốt nghiệp [$F(3,3923) = 36,082, p < 0,001$]. Kiểm tra Tukey HSD sau kiểm định cho thấy thí sinh tốt nghiệp đại học gần đây từ năm 2018 đến năm 2022 có số câu trả lời đúng cao hơn đáng kể so với những thí sinh tốt nghiệp trước năm 2014 ($p < 0,001$).

BÀN LUẬN

Nghiên cứu hiện tại sử dụng phương pháp phân tích câu hỏi hồi cứu để khám phá mối quan hệ giữa độ khó của câu hỏi logit và chỉ số phân biệt r_{pbis} trong bài kiểm tra đầu vào chương trình bác sĩ chuyên khoa cấp I, dựa trên môn thi Sinh lý học.

Trái ngược với giả định rằng các câu hỏi dễ không có khả năng phân biệt giữa những người học đạt điểm cao và thấp, nghiên cứu của chúng tôi cho thấy rằng độ khó của câu hỏi không thể xác định tiềm năng phân cách năng lực thí sinh. Ngoài ra, những phát hiện cho thấy rằng các yếu tố bên ngoài độ khó, chẳng hạn như cơ sở đào

tạo ở bậc đại học, xếp loại tốt nghiệp đại học, thời gian tốt nghiệp đại học đến khi đăng ký dự thi bậc đào tạo sau đại học, có thể đóng vai trò quan trọng trong việc ảnh hưởng đến độ phân cách. Điều này nhấn mạnh tầm quan trọng của việc xem xét đa yếu tố khi đánh giá khả năng phân biệt của các câu hỏi trắc nghiệm. Nghiên cứu của chúng tôi góp phần bổ sung minh chứng cho thấy độ khó và khả năng phân biệt của câu hỏi không nên được coi là các đặc điểm loại trừ lẫn nhau nói chung, và trong đề kiểm tra môn Sinh lý học nói riêng.

Một số đặc điểm (năng lực) của thí sinh có thể góp phần tạo ra hiện tượng các câu hỏi được phân loại là dễ nhưng có khả năng biệt cao.

Hơn nữa, chúng tôi thấy sự chênh lệch đáng kể về kết quả thi của thí sinh ở các mức độ học lực, năm tốt nghiệp và trường tốt nghiệp đại học khác nhau. Có thể nghĩ đến một trong những nguyên nhân là do kiến thức và kỹ năng của thí sinh đạt xếp loại tốt nghiệp càng tốt thì sẽ có năng lực càng cao hơn và số câu làm đúng của cũng sẽ nhiều hơn. Bên cạnh đó, chúng tôi cũng nghĩ đến những trường hợp thí sinh tốt nghiệp lâu hơn, có thể kiến thức đã quên nhiều hơn nên kết quả thi không tốt bằng so với các thí sinh mới tốt nghiệp gần đây. Ngoài ra, yếu tố trường tốt nghiệp đại học cũng góp phần không nhỏ tạo nên sự khác biệt năng lực của thí sinh. Sự chênh lệch này có khả năng làm cho sức mạnh phân biệt của các câu hỏi trắc nghiệm (MCQs) trở nên cao hơn, ngay cả khi các câu hỏi dễ. Hiện tượng này xảy ra bởi vì khi có sự chênh lệch đáng kể về năng lực giữa các thí sinh, các câu hỏi có thể phân biệt hiệu quả giữa những thí sinh có năng lực trên trung bình và kém dưới ngưỡng. Trong trường hợp như vậy, ngay cả những câu hỏi dễ cũng có thể có độ phân cách tốt [4,5].

Như vậy, các câu hỏi dễ có thể có sức mạnh phân biệt cao là do chênh lệch trình độ của thí sinh. Điều quan trọng là phải nhận ra rằng chỉ riêng mức độ khó thì không xác định khả năng của một câu hỏi trong việc phân biệt giữa các năng lực. do đó, chúng tôi cần nghiên cứu thêm với phân tích câu theo Lý thuyết ứng đáp câu hỏi, trong đó độ khó câu hỏi được đo lường trên thang logit. Điều này có thể giúp thông tin thêm cho việc phát triển các chiến lược đánh giá đo lường chính xác năng lực của người học và thúc đẩy việc đánh giá công bằng.

KẾT LUẬN

Nghiên cứu của chúng tôi ghi nhận các câu hỏi dễ trong bài kiểm tra môn Sinh lý học có thể có độ phân cách tốt. Những phát hiện này nhấn mạnh tầm quan trọng của việc xem xét nhiều yếu tố vượt ra ngoài độ khó của câu hỏi khi đánh giá chất lượng câu hỏi và thúc đẩy việc đánh giá chính xác. Cuối cùng, một cách tiếp cận cân bằng kết hợp cả độ khó và khả năng phân biệt của câu hỏi giúp việc đánh giá hợp lệ và ý nghĩa

hơn về năng lực của học viên trong lĩnh vực Sinh lý học hay bài thi phối hợp các lĩnh vực khác.

TÀI LIỆU THAM KHẢO

1. De Caldas, Francisco José. Classical test theory and item response theory: Two understandings of one high-stakes performance exam. *Colombian Applied Linguistics Journal*. 2014; 16(2):167-184.
2. Azzopardi, M., & Azzopardi, C. (2019). Relationship between item difficulty level and item discrimination in biology final examinations. *Education and New Developments*, 2019.
3. Karadag, N., & Sahin, M. D. (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and Distance Learning Context. *Turkish Online Journal of Educational Technology-TOJET*, 15(4), 16-24.
4. Pimentel, J. L., & Villaruz, M. L. A. (2020). Comparison of Item Difficulty Estimates in a Basic Statistics Test using ltm and CTT Software Packages in R. *International Journal of Advanced Computer Science and Applications*, 11(3).
5. Fatima, S., Ahmed, S. T., Hashmat, S., Iftikhar, A., & Hashmi, T. A. (2021). The relationship of difficulty and discrimination indices with physiology teaching and learning. *Pak J Med Dent*, 10(2), 109-13.
6. Barker, B. A., Donovan, N. J., Schubert, A. D., & Walker, E. A. (2017). Using Rasch analysis to examine the item-level psychometrics of the Infant-Toddler Meaningful Auditory Integration Scales. *Speech, Language and hearing*, 20(3), 130-143.