

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC GIÁO DỤC

BÙI THỊ KIM PHƯỢNG

**XÂY DỰNG BÀI KIỂM TRA
THÍCH ỨNG BẰNG MÁY TÍNH ĐỂ
ĐÁNH GIÁ KIẾN THỨC TỪ VỰNG
TIẾP NHẬN TIẾNG ANH**

LUẬN ÁN TIẾN SĨ ĐO LƯỜNG VÀ ĐÁNH GIÁ TRONG GIÁO DỤC

Mã số: 9140115

HÀ NỘI – 2024

ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC GIÁO DỤC

BÙI THỊ KIM PHƯƠNG

**XÂY DỰNG BÀI KIỂM TRA
THÍCH ỨNG BẰNG MÁY TÍNH ĐỂ
ĐÁNH GIÁ KIẾN THỨC TỪ VỰNG
TIẾP NHẬN TIẾNG ANH**

**LUẬN ÁN TIẾN SĨ ĐO LƯỜNG VÀ ĐÁNH GIÁ TRONG GIÁO DỤC
CHUYÊN NGÀNH: ĐO LƯỜNG VÀ ĐÁNH GIÁ TRONG GIÁO DỤC**

Mã số: 9140115

**Người hướng dẫn khoa học: 1. GS.TS. NGUYỄN QUÝ THANH
2. PGS.TS. LÊ THÁI HÙNG**

NGHIÊN CỨU SINH

CÁN BỘ HƯỚNG DẪN

XÁC NHẬN CỦA ĐƠN VỊ ĐÀO TẠO

HÀ NỘI – 2024

LỜI CAM ĐOAN

Tôi xin cam đoan luận án tiến sĩ “**Xây dựng bài kiểm tra thích ứng bằng máy tính để đánh giá kiến thức từ vựng tiếp nhận tiếng Anh**” là công trình nghiên cứu của chính bản thân tôi.

Trong quá trình thực hiện luận án, tôi đã tuân thủ nghiêm túc các quy tắc đạo đức nghiên cứu; các nội dung trình bày trong luận án là trung thực, và không sao chép từ bất kỳ một nguồn nào và dưới bất kỳ hình thức nào. Việc tham khảo các nguồn tài liệu (nếu có) đã được thực hiện trích dẫn và ghi nguồn tài liệu tham khảo đúng quy định.

Tôi xin hoàn toàn chịu trách nhiệm về tính trung thực của các nội dung trong luận án của mình.

Hà Nội, ngày 15 tháng 07 năm 2024

Tác giả luận án

Bùi Thị Kim Phượng

LỜI CẢM ƠN

Để có thể hoàn thành được luận án tiến sĩ này, tôi đã nhận được sự hỗ trợ và giúp đỡ từ gia đình, các thầy cô, bạn bè, đồng nghiệp và các em sinh viên.

Trước hết, tôi xin bày tỏ lòng biết ơn sâu sắc tới GS.TS. Nguyễn Quý Thanh và PGS.TS. Lê Thái Hưng đã tận tình hướng dẫn, chỉ bảo tôi trong suốt quá trình học tập cũng như thực hiện luận án.

Tôi xin trân trọng cảm ơn Ban giám hiệu trường Đại học Giáo dục - ĐHQGHN, Ban chủ nhiệm và các thầy cô Khoa Quản trị chất lượng, trường ĐHGĐ – Đại học Quốc gia Hà Nội đã hướng dẫn, giúp đỡ, tạo điều kiện thuận lợi cho tôi trong suốt thời gian học tập và nghiên cứu. Xin gửi lời cảm ơn nhiệt thành gửi tới PGS.TS. Nguyễn Thúy Nga, PGS.TS. Vũ Trọng Lương, TS. Tăng Thị Thùy, TS. Trần Thị Thu Hương, TS. Trần Xuân Quang và biết bao các thầy cô đã giúp tôi củng cố kiến thức và cho tôi những lời khuyên quý báu trong thời gian thực hiện luận án.

Tôi xin chân thành cảm ơn Ban lãnh đạo của Khoa Ngoại ngữ - Đại học Bách khoa Hà Nội, nơi tôi đang công tác cũng các đồng nghiệp đã luôn tin tưởng, ủng hộ tôi ngay từ những ngày đầu thực hiện luận án. Không có được sự hỗ trợ này cùng sự nhiệt tình tham gia của các em sinh viên, tôi sẽ không thể nào hoàn thành luận án.

Cuối cùng, tôi dành tất cả sự yêu thương và lời cảm ơn tận đáy lòng tới gia đình của tôi, những người thân yêu đã luôn động viên, khích lệ, ủng hộ tôi trong suốt quá trình học tập và thực hiện luận án.

Một lần nữa, tôi xin trân trọng cảm ơn!

DANH MỤC CHỮ VIẾT TẮT

Cụm từ viết tắt	Cụm từ đầy đủ (nghĩa tiếng Việt)
CALT	: Computerized Adaptive Language Testing Kiểm tra ngôn ngữ thích ứng trên máy tính
CAT	: Computerized Adaptive Testing Kiểm tra thích ứng bằng/trên máy tính
CATSS	: Computer adaptive test of size and strength Bài kiểm tra từ vựng thích ứng bằng máy tính đánh giá độ rộng và độ sâu
CAT-WPLT	: Computerized Adaptive Testing – Word Part Levels Test Bài kiểm tra thích ứng về mức độ hiểu biết thành tố từ
CNTT	: Công nghệ thông tin
ĐHGD	: Trường Đại học Giáo dục
ĐHQGHN	: Đại học Quốc gia Hà Nội
IRT	: Item Response Theory Lý thuyết hồi đáp câu hỏi
NGSL	: New General Service List Danh sách từ vựng tiếng Anh thông dụng mới
NGSLT	: New General Service List Test Bài kiểm tra danh sách từ vựng tiếng Anh thông dụng mới
TS	: thí sinh
VLT	: Vocabulary Levels Test Bài kiểm tra cấp độ từ vựng
VST	: Vocabulary Size Test Bài kiểm tra độ rộng từ vựng
UEd-ALS	: University of Education – Adaptive Learning System Hệ thống học tập thích ứng của trường ĐHQGHN
UEd-CAT	: University of Education – Computerized Adaptive Testing Hệ thống trắc nghiệm thích ứng của trường ĐHQGHN

DANH MỤC HÌNH

Hình 1.1: Biểu đồ CAT (Thompson & Weiss, 2011)	14
Hình 1.2: Quy trình xây dựng ngân hàng câu hỏi.....	16
Hình 1.3: Thang đo kiến thức từ vựng (Paribakht & Welshe, 1997).....	24
Hình 1.4: Các khía cạnh của kiến thức từ vựng (Nation, 2013)	25
Hình 1.5: Quy trình xây dựng đề kiểm tra (Bachman & Palmer, 1996).....	29
Hình 1.6: Khung kiểm tra từ vựng (Read & Chapelle, 2001).....	33
Hình 1.7: Ví dụ câu hỏi trong VLT.....	51
Hình 1.8: Ví dụ câu hỏi trong New VLT	53
Hình 1.9: Ví dụ câu hỏi VST	55
Hình 1.10: Mô hình nghiên cứu	63
Hình 2.1: Quy trình nghiên cứu	66
Hình 2.2: Thứ tự sử dụng các phương pháp nghiên cứu.....	67
Hình 2.3: Hướng dẫn làm bài trên hệ thống.....	77
Hình 2.4: Quy trình phát triển hệ thống trắc nghiệm thích ứng.....	80
Hình 2.5: Các bước của một bài trắc nghiệm thích ứng	86
Hình 2.6: Tính năng của UEd-CAT	87
Hình 2.7: Câu hỏi ví dụ trong NGSLT.....	88
Hình 2.8: Câu hỏi ví dụ của NGSLT	89
Hình 2.9: Câu hỏi ví dụ trong NGSLT song ngữ tiếng Anh và tiếng Việt	92
Hình 2.10: Bản đồ phân bố năng lực và độ khó của đề mẫu	93
Hình 2.11: Đường cong đặc trưng của câu hỏi 66	96
Hình 2.12: Xác nhận của người tham gia nghiên cứu.....	98
Hình 3.1: Bản đồ phân bố năng lực và độ khó của Đề 7	105
Hình 3.2: Kết quả phân tích Conquest của Đề 1	106
Hình 3.3: Đường cong đặc trưng của câu hỏi 20 – Đề 4	108

Hình 3.4: Kết quả phân tích Conquest của Đề 6.....	109
Hình 3.5: Sơ đồ neo giữa các đề	113
Hình 3.6: Độ khó câu hỏi thi trước và sau khi cân bằng.....	115
Hình 3.7: Độ khó của ngân hàng câu hỏi chuẩn hóa	115
Hình 3.8: Thời gian làm bài trên hệ thống UEd-CAT	117
Hình 3.9: Lộ trình thích ứng trong bài kiểm tra của thí sinh HONG.....	118
Hình 3.10: Sai số chuẩn của phép ước lượng năng lực cập nhật sau từng câu hỏi trong bài làm của thí sinh HONG	119
Hình 3.11: Lộ trình thích ứng lượt làm bài số 1 của thí sinh DAN	121
Hình 3.12: Lộ trình thích ứng lượt làm bài số 2 của thí sinh DAN	121
Hình 3.13: Lộ trình thích ứng lượt làm bài số 3 của thí sinh DAN	122
Hình 3.14: Điểm của thí sinh với bài kiểm tra 20 câu hỏi	123
Hình 3.15: Lộ trình thích ứng trong bài trắc nghiệm HONG thực hiện	124
Hình 3.16: Lộ trình thích ứng trong bài trắc nghiệm MDUC thực hiện.....	125
Hình 3.17: Lộ trình thích ứng trong bài trắc nghiệm LINH thực hiện	125
Hình 3.18: Biểu đồ phân tán tỉ lệ trả lời chính xác và điểm bài kiểm tra thích ứng.....	127
Hình 3.19: Kết quả làm bài kiểm tra cố định của 98 thí sinh	128
Hình 3.20: Biểu đồ phân tán điểm số trong bài kiểm tra cố định và bài kiểm tra thích ứng	129
Hình 3.21: Giá trị trung bình mức độ đồng ý với các nhận định về	132
Hình 3.22: Mức độ mong muốn của người tham gia khảo sát về.....	134

DANH MỤC BẢNG

Bảng 1.1: Các mô hình IRT (Lâm Quang Thiệp, 2010; Carlson, 2020)	12
Bảng 1.2: Khung xây dựng CAT (Thompson & Weiss, 2011)	18
Bảng 1.3: Các yếu tố xác định mục đích kiểm tra	31
Bảng 1.4: Từ vựng trong Nội dung dạy học các cấp của Chương trình giáo dục phổ thông môn Tiếng Anh 2018	36
Bảng 1.5: Đặc tả về từ vựng theo các bậc năng lực ngôn ngữ.....	38
Bảng 1.6: Định dạng câu hỏi trong CATSS.....	57
Bảng 2.1: Thông tin mẫu của phương pháp chuyên gia	75
Bảng 2.2: Cỡ mẫu thử nghiệm theo đề	76
Bảng 2.3: Tổng hợp số liệu sinh viên thử nghiệm trên hệ thống.....	77
Bảng 2.4: Thông tin của người tham gia khảo sát	78
Bảng 2.5: Thông tin của người tham gia phỏng vấn.....	79
Bảng 2.6: Bảng đặc tả bài kiểm tra song ngữ đánh giá từ vựng tiếp nhận tiếng Anh.....	90
Bảng 2.7: Hệ số Alpha và hệ số tin cậy độc lập của đề mẫu	92
Bảng 2.8: Các câu hỏi không phù hợp với mô hình của đề mẫu	94
Bảng 2. 9: Kết quả phân tích của câu hỏi 66	95
Bảng 3.1: Số câu hỏi thô theo mức độ tần suất.....	101
Bảng 3.2: Thông tin thẩm định đề của nhóm chuyên gia	101
Bảng 3.3: Tổng hợp đánh giá của nhóm chuyên gia.....	102
Bảng 3.4: Ví dụ câu hỏi chỉnh sửa theo ý kiến chuyên gia.....	103
Bảng 3.5: Cỡ mẫu được chọn phân tích của bảy bài kiểm tra thử nghiệm...	103
Bảng 3.6: Độ tin cậy theo đề.....	104
Bảng 3.7: Các câu hỏi không phù hợp với mô hình của Đề 1	107
Bảng 3.8: Kết quả phân tích của câu hỏi 20 Đề 4.....	107

Bảng 3.9: Tổng hợp số câu hỏi cần chỉnh sửa	109
Bảng 3.10: Các câu hỏi cần chỉnh sửa của Đề 6	110
Bảng 3.11: Thiết kế câu hỏi neo giữa bảy đề sau khi phân tích Conquest ...	113
Bảng 3.12: Hệ số cân bằng bảy đề thử nghiệm.....	114
Bảng 3.13: Các gói câu hỏi	119
Bảng 3.14: Ba lượt làm bài của thí sinh DAN	120
Bảng 3.15: Thông số của các bài kiểm tra 20 câu hỏi ba thí sinh thực hiện.	124
Bảng 3.16: Kết quả làm bài trong các lượt làm bài khác nhau	126
Bảng 3.17: Phân tích tương quan Pearson giữa tỉ lệ trả lời chính xác và điểm bài kiểm tra thích ứng.....	127
Bảng 3.18: Các trường hợp bị loại trừ	129
Bảng 3.19: Nhận thức của học sinh về đặc điểm bài kiểm tra thích ứng	131

MỤC LỤC

DANH MỤC CHỮ VIẾT TẮT	iii
DANH MỤC HÌNH	iv
DANH MỤC BẢNG.....	vi
MỤC LỤC.....	viii
MỞ ĐẦU	1
1. Đặt vấn đề.....	1
2. Mục đích và nhiệm vụ nghiên cứu.....	3
2.1. Mục đích nghiên cứu.....	3
2.2. Nhiệm vụ nghiên cứu	4
3. Khách thể và đối tượng nghiên cứu	4
4. Phạm vi và giới hạn nghiên cứu.....	4
5. Câu hỏi nghiên cứu	5
6. Phương pháp nghiên cứu.....	5
7. Đóng góp khoa học của luận án.....	6
8. Cấu trúc của luận án	8
CHƯƠNG 1: CƠ SỞ LÝ LUẬN VÀ TỔNG QUAN NGHIÊN CỨU.....	9
1.1. Cơ sở lý luận	9
1.1.1. Lý thuyết khảo thí hiện đại.....	9
1.1.2. Lý luận về trắc nghiệm thích ứng bằng máy tính.....	14
1.1.3. Lý luận về đánh giá kiến thức từ vựng tiếp nhận tiếng Anh	21
1.2. Tổng quan nghiên cứu.....	39
1.2.1. Các nghiên cứu về kiểm tra thích ứng trong đào tạo ngôn ngữ	39
1.2.2. Các nghiên cứu về đánh giá kiến thức từ vựng tiếp nhận tiếng Anh	49
1.2.3. Khoảng trống nghiên cứu.....	61
1.3. Đề xuất mô hình nghiên cứu	63
1.4. Kết chương.....	64
CHƯƠNG 2: PHƯƠNG PHÁP LUẬN NGHIÊN CỨU.....	65
2.1. Quy trình nghiên cứu	65
2.2. Phương pháp nghiên cứu.....	67
2.2.1. Phương pháp nghiên cứu định lượng	68
2.2.2. Phương pháp nghiên cứu định tính.....	72

2.3. Quá trình lấy mẫu.....	74
2.3.1. Mẫu của phương pháp chuyên gia.....	74
2.3.2. Mẫu tham gia thử nghiệm.....	75
2.3.3. Mẫu tham gia khảo sát và phỏng vấn.....	78
2.4. Công cụ nghiên cứu.....	80
2.4.1. Hệ thống UEd-CAT.....	80
2.4.2. Bài trắc nghiệm song ngữ đánh giá từ vựng tiếp nhận tiếng Anh.....	87
2.4.3. Bảng câu hỏi xin ý kiến chuyên gia.....	96
2.4.4. Bảng câu hỏi khảo sát.....	96
2.4.5. Bộ câu hỏi phỏng vấn.....	97
2.5. Các vấn đề về đạo đức nghiên cứu.....	98
2.6. Kết chương.....	99
CHƯƠNG 3: KẾT QUẢ NGHIÊN CỨU VÀ BÀN LUẬN.....	100
3.1. Chuẩn hóa ngân hàng câu hỏi trắc nghiệm thích ứng từ vựng tiếp nhận tiếng Anh.....	100
3.1.1. Biên soạn và chỉnh sửa bộ câu hỏi thô.....	100
3.1.2. Đánh giá độ tin cậy của các đề thử nghiệm.....	103
3.1.3. Loại các câu hỏi không phù hợp với mô hình.....	106
3.1.4. Phân loại và chỉnh sửa câu hỏi.....	107
3.1.5. Cân bằng đề và chuẩn hóa ngân hàng câu hỏi.....	112
3.2. Đánh giá bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh.....	116
3.2.1. Quá trình làm bài của thí sinh trên hệ thống UEd-CAT.....	116
3.2.2. Kết quả làm bài của thí sinh trên hệ thống.....	122
3.2.3. Phản hồi của thí sinh thực hiện trắc nghiệm thích ứng.....	130
3.3. Bàn luận.....	141
3.4. Kết chương.....	145
KẾT LUẬN.....	146
1. Tóm lược kết quả nghiên cứu của luận án.....	146
2. Đóng góp của luận án.....	147
3. Khuyến nghị.....	150
3.1. Khuyến nghị với người học.....	150
3.2. Khuyến nghị với giáo viên và các cơ sở đào tạo.....	151

3.3. Khuyến nghị với nhóm chuyên gia phát triển hệ thống.....	151
3.4. Khuyến nghị với các nhà nghiên cứu.....	152
3.5. Khuyến nghị với các cơ quan quản lý giáo dục.....	152
4. Hạn chế của luận án và đề xuất nghiên cứu tiếp theo.....	153
DANH MỤC CÔNG TRÌNH NGHIÊN CỨU.....	155
TÀI LIỆU THAM KHẢO.....	156
Phụ lục 1: Bảng câu hỏi xin ý kiến chuyên gia.....	175
Phụ lục 2: Bảng câu hỏi khảo sát sinh viên.....	178
Phụ lục 3: Bộ câu hỏi phỏng vấn sinh viên.....	182
Phụ lục 4: Đề kiểm tra song ngữ từ vựng tiếp nhận tiếng Anh.....	184
Phụ lục 5: Kết quả đánh giá định tính 7 đề kiểm tra từ các chuyên gia.....	191
Phụ lục 6: Kết quả phân tích sự phù hợp với mô hình của 7 đề thử nghiệm.....	193
Phụ lục 7: Bản đồ phân bố năng lực và độ khó của 7 đề thử nghiệm.....	207
Phụ lục 8: Ngân hàng câu hỏi đã chuẩn hóa nhập trên hệ thống UEd-CAT.....	214
Phụ lục 9: Báo cáo thử nghiệm trên hệ thống UEd-CAT.....	215
Phụ lục 10: Nội dung phỏng vấn.....	221

MỞ ĐẦU

1. Đặt vấn đề

Ứng dụng công nghệ thông tin vào giáo dục là một trong 9 nhiệm vụ trong giai đoạn 2016-2020 và định hướng 2025 của ngành giáo dục Việt Nam. Đặc biệt hơn, kỷ nguyên 4.0 của chuyển đổi kỹ thuật số đã tác động đến tất cả các khía cạnh của giáo dục và thúc đẩy các phương pháp kiểm tra đánh giá với nhiều đổi mới. Trong lĩnh vực đánh giá ngôn ngữ, việc ứng dụng CNTT đã trở nên phổ biến hơn ở tất cả các gia đình và trường học, do đó tạo điều kiện thuận lợi cho một sáng kiến kiểm tra đánh giá hiệu quả hơn - một hệ thống kiểm tra ngôn ngữ thích ứng trên máy tính. Trên thế giới, ngày càng có nhiều bài kiểm tra ngôn ngữ thích ứng trên máy tính được phát triển và nhận được phản hồi tích cực; trong khi đó, tại Việt Nam chưa có bài kiểm tra ngôn ngữ thích ứng nào được phát triển và công bố.

Trong bối cảnh giáo dục ở Việt Nam, hệ thống trắc nghiệm thích ứng được xem là tiên phong và duy nhất đến thời điểm hiện tại là UEd-CAT. Hệ thống được xây dựng và phát triển bởi trường Đại học Giáo Dục – Đại học Quốc gia Hà Nội, và đã công bố những kết quả rất tích cực trong việc kiểm tra đánh giá về toán và đọc hiểu tiếng Việt, tạo động lực cho việc phát triển các bài kiểm tra thích ứng bằng máy tính hướng tới việc đánh giá ngôn ngữ trong thời gian tới.

Với lĩnh vực kiểm tra ngôn ngữ, kiểm tra kiến thức từ vựng có ý nghĩa quan trọng trong đào tạo ngôn ngữ với cả người dạy và người học. Với người học, việc kiểm tra kiến thức từ vựng giúp xác định trình độ thông thạo ngôn ngữ của người học vì kiến thức từ vựng đóng vai trò nền móng cho tất cả các hoạt động sử dụng ngôn ngữ (Schmitt và cộng sự, 2017). Lĩnh hội được một

lượng kiến thức từ vựng là một trong những điều kiện tiên quyết quan trọng để học ngôn ngữ thành công. Với người dạy, việc có những ước tính đáng tin cậy về kiến thức từ vựng cho phép giáo viên cung cấp tài liệu phù hợp cho nhu cầu của người học, đánh giá hiệu quả của quá trình học và đặt ra các mục tiêu phù hợp để người học có thể phát triển kỹ năng và năng lực ngôn ngữ của mình (Nation, 2013). Đối với mục đích nghiên cứu, kiến thức từ vựng trở thành một yếu tố dự báo mạnh mẽ về trình độ ngôn ngữ của người học và thậm chí cả thành tích học tập của họ (Lin & Morrison, 2010). Ở chiều hướng ngược lại, năng lực từ vựng của người học có xu hướng cải thiện khi trình độ ngôn ngữ của họ phát triển (Zareva và cộng sự, 2005), hay quá trình áp dụng bốn kỹ năng ngôn ngữ là đọc, nghe, nói và viết trong giao tiếp hỗ trợ việc thu nhận các từ mới học vào bộ nhớ (Laufer, 2013). Ngoài ra, các bài kiểm tra từ vựng có thể được sử dụng để đánh giá tác động của trải nghiệm học tập đối với quá trình phát triển từ vựng cũng như để đo lường mức độ phát triển từ vựng (Stoeckel & Bennett, 2015). Đã có nhiều bài kiểm tra từ vựng được thiết kế và sử dụng để đánh giá các khía cạnh khác nhau về kiến thức từ vựng của người học, tuy nhiên các nhà nghiên cứu hàng đầu vẫn có những tranh luận về điểm mạnh yếu và đề xuất các hướng phát triển các bài trắc nghiệm từ vựng mới áp dụng lý thuyết khảo thí hiện đại cũng như những thành tựu công nghệ mới để mang lại lợi ích cho các bên liên quan (Schmitt và cộng sự, 2020).

Trong bối cảnh dạy và học ngoại ngữ Việt Nam, từ vựng luôn có được xem trọng trong chương trình giảng dạy tiếng Anh ở Việt Nam. Trong Chương trình giáo dục phổ thông môn Tiếng Anh được ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo, học sinh sau khi hoàn thành chương trình phổ thông, cần có số lượng từ vựng là khoảng 2500 từ. Tuy nhiên, theo kết quả của một số lượng không nhiều các nghiên cứu gần đây kiểm tra từ vựng của người học tiếng Anh

của Việt Nam, học sinh phổ thông và sinh viên đại học có lượng từ vựng rất hạn chế, không đạt được yêu cầu về lượng từ này (Vu & Peters, 2021). Việc có thêm những công cụ đánh giá kiến thức từ vựng tiếng Anh của người học cũng được đặt ra như một nhiệm vụ quan trọng để hỗ trợ người dạy, người học cũng như những nhà nghiên cứu trong lĩnh vực dạy và học tiếng Anh tại Việt Nam.

Xuất phát từ những nhận định trên, đề tài “**Xây dựng bài kiểm tra thích ứng bằng máy tính để đánh giá kiến thức từ vựng tiếp nhận tiếng Anh**” được lựa chọn làm đề tài nghiên cứu thuộc chuyên ngành Đo lường và đánh giá trong giáo dục. Việc phát triển bài trắc nghiệm thích ứng trên máy tính đánh giá từ vựng tiếng Anh, hướng tới việc ứng dụng và nâng cao hiệu quả của quá trình dạy và học là phù hợp với xu hướng phát triển trong giáo dục để đáp ứng yêu cầu đổi mới trong kỷ nguyên chuyển đổi số, hứa hẹn mang lại những đóng góp có giá trị trong lĩnh vực đào tạo ngôn ngữ cũng như trong lĩnh vực đo lường và đánh giá trong giáo dục tại Việt Nam.

2. Mục đích và nhiệm vụ nghiên cứu

2.1. Mục đích nghiên cứu

Luận án được thực hiện với mục đích xây dựng bài kiểm tra thích ứng bằng máy tính để đánh giá từ vựng tiếp nhận tiếng Anh dành cho người học ngoại ngữ tiếng Anh tại Việt Nam. Với việc sử dụng các thuật toán sẵn có của hệ thống trắc nghiệm thích ứng UEd-CAT, luận án tập trung vào tìm hiểu các thuật toán để thiết kế ngân hàng câu hỏi kiểm tra từ vựng tiếp nhận tiếng Anh đáp ứng yêu cầu của hệ thống, từ đó tiến hành xây dựng, thử nghiệm và đánh giá bài trắc nghiệm thích ứng đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của người học ngoại ngữ tiếng Anh tại Việt Nam.

2.2. Nhiệm vụ nghiên cứu

Để thực hiện được mục đích nghiên cứu, luận án triển khai những nhiệm vụ nghiên cứu cụ thể như sau:

- (1) Xây dựng và chuẩn hóa ngân hàng câu hỏi trắc nghiệm đánh giá kiến thức từ vựng tiếp nhận tiếng Anh đáp ứng các yêu cầu của hệ thống trắc nghiệm thích ứng UEd-CAT.
- (2) Thiết kế, thử nghiệm và đánh giá bài kiểm tra thích ứng bằng máy tính đánh giá kiến thức từ vựng tiếp nhận tiếng Anh.

3. Khách thể và đối tượng nghiên cứu

- **Khách thể nghiên cứu:** người học ngoại ngữ tiếng Anh ở Việt Nam trong hoạt động kiểm tra đánh giá kiến thức từ vựng tiếp nhận tiếng Anh

- **Đối tượng nghiên cứu:** bài kiểm tra thích ứng bằng máy tính đánh giá kiến thức từ vựng tiếp nhận tiếng Anh

4. Phạm vi và giới hạn nghiên cứu

- **Phạm vi nghiên cứu:** Luận án tập trung vào việc xây dựng và chuẩn hóa đề trắc nghiệm thích ứng để đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của người học ngoại ngữ tiếng Anh ở Việt Nam.

- **Giới hạn nghiên cứu:** Về thời gian thực hiện, với quy mô của luận án, thử nghiệm, khảo sát và phỏng vấn được lên kế hoạch và thực hiện trong khoảng thời gian từ tháng 12/2020 đến tháng 12/2023. Về đối tượng tham gia nghiên cứu, luận án được thực hiện với sinh viên các chuyên ngành kỹ thuật tại một trường đại học tại Hà Nội, một nhóm đối tượng người học ngoại ngữ tiếng Anh ở Việt Nam. Về bối cảnh thực hiện nghiên cứu, luận án sử dụng hệ thống trắc

nghiệm thích ứng của trường ĐHGD - ĐHQGHN với sự cho phép của nhóm chuyên gia phát triển hệ thống.

5. Câu hỏi nghiên cứu

Căn cứ vào mục đích, nhiệm vụ và đối tượng nghiên cứu, luận án được thực hiện để trả lời hai câu hỏi nghiên cứu sau:

Câu hỏi 1: Ngân hàng câu hỏi trắc nghiệm đánh giá kiến thức từ vựng tiếp nhận tiếng Anh được xây dựng và chuẩn hóa như thế nào?

Câu hỏi 2: Bài kiểm tra thích ứng bằng máy tính được thiết kế thực hiện việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của người học ngoại ngữ tiếng Anh ở Việt Nam như thế nào?

6. Phương pháp nghiên cứu

Để đạt được mục đích của nghiên cứu, luận án sử dụng các phương pháp nghiên cứu khoa học kết hợp định tính và định lượng như sau:

Phương pháp nghiên cứu định tính

- Phương pháp chuyên gia nhằm thu thập các ý kiến của những người có kinh nghiệm, có kiến thức chuyên sâu trong lĩnh vực nghiên cứu có liên quan.

- Phương pháp phỏng vấn để tìm hiểu sâu quan điểm của thí sinh, cung cấp thêm góc nhìn về bài kiểm tra thích ứng được thiết kế.

- Phương pháp phân tích nội dung để tổng thuật và nghiên cứu các quan điểm, công trình nghiên cứu có liên quan ở trong và ngoài nước làm cơ sở cho việc xây dựng khung lí thuyết của đề tài, định hướng cho nghiên cứu thực tiễn cũng như phân tích nhận thức của người tham gia khảo sát và phỏng vấn.

Phương pháp nghiên cứu định lượng

- Phương pháp thử nghiệm nhằm chuẩn hóa ngân hàng câu hỏi và đánh giá bài kiểm tra thích ứng được thiết kế.
- Phương pháp khảo sát để tìm hiểu nhận thức của thí sinh về bài kiểm tra thích ứng được thiết kế.
- Phương pháp xử lý số liệu toán học và thống kê với các phần mềm Excel, SPSS, Conquest, R để phân tích số liệu.

7. Đóng góp khoa học của luận án

Đóng góp về lý luận

Luận án là công trình nghiên cứu khoa học có hệ thống, logic, và chặt chẽ dựa trên cơ sở lý thuyết được phát triển bởi các học giả và nhà nghiên cứu liên quan về kiểm tra từ vựng và trắc nghiệm thích ứng bằng máy tính.

- Luận án hệ thống hóa vấn đề lý luận về kiểm tra từ vựng và việc áp dụng trắc nghiệm thích ứng trong kiểm tra từ vựng tiếng Anh, cụ thể là đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của đối tượng người học tiếng Anh ở Việt Nam.

- Luận án là một trong những nghiên cứu quy mô đầu tiên về trắc nghiệm thích ứng bằng máy tính áp dụng trong đào tạo ngôn ngữ ở Việt Nam, đóng góp bằng chứng xác thực về việc áp dụng lý thuyết hồi đáp và phương pháp cân bằng trong xây dựng và chuẩn hóa ngân hàng câu hỏi trong lĩnh vực đo lường và đánh giá trong giáo dục.

- Luận án hứa hẹn đóng góp vào lĩnh vực kiểm tra đánh giá ngôn ngữ những giá trị lý luận có ý nghĩa hướng tới đối tượng người học tiếng Anh ở

Việt Nam, từ đó mang lại những đóng góp tích cực vào việc áp dụng CNTT trong đo lường và đánh giá trong giáo dục cũng như lĩnh vực dạy và học ngoại ngữ tiếng Anh ở Việt Nam.

Đóng góp về thực tiễn

- Luận án xây dựng, thử nghiệm và đánh giá hiệu quả của một công cụ kiểm tra kiến thức từ vựng tiếp nhận tiếng Anh hướng tới người học ngoại ngữ tiếng Anh tại Việt Nam. Với việc áp dụng trắc nghiệm thích ứng bằng máy tính, công cụ hứa hẹn có những tính năng vượt trội, mang lại tính chính xác và hiệu quả đánh giá cao khi so với các bài kiểm tra cố định đã được phát triển trước đây.

- Luận án cung cấp những bằng chứng xác thực để khẳng định tính khả thi của việc áp dụng trắc nghiệm thích ứng bằng máy tính trong kiểm tra từ vựng. Cùng với một số lượng ít các nghiên cứu được thực hiện về trắc nghiệm thích ứng bằng máy tính ở Việt Nam, luận án mở đường cho các nghiên cứu trong tương lai về việc áp dụng trắc nghiệm thích ứng trong kiểm tra và đánh giá với các nội dung và mục đích khác.

- Luận án thu nhận những kết quả đánh giá từ quá trình thử nghiệm cũng như từ góc nhìn của các thí sinh có trải nghiệm trực tiếp, vì vậy có thể cung cấp những ý tưởng cũng như nhận định có giá trị và đáng tin cậy về việc áp dụng trắc nghiệm thích ứng trong kiểm tra đánh giá cũng như trong quá trình dạy và học, để có thể mang lợi ích cho cả người dạy và người học, cũng như các nhà nghiên cứu hay nhóm phát triển hệ thống trắc nghiệm thích ứng.

8. Cấu trúc của luận án

Luận án gồm có ba phần chính: mở đầu, nội dung nghiên cứu và kết luận.

Phần Mở đầu là phần giới thiệu tổng thể luận án, gồm có phần đặt vấn đề, mục đích và nhiệm vụ nghiên cứu, khách thể và đối tượng nghiên cứu, phạm vi và giới hạn nghiên cứu, câu hỏi nghiên cứu, phương pháp nghiên cứu, đóng góp của luận án về lý luận và thực tiễn, cũng như cấu trúc của luận án.

Phần Nội dung nghiên cứu có 3 chương chính:

- Chương 1: Cơ sở lý luận và tổng quan nghiên cứu
- Chương 2: Thiết kế nghiên cứu
- Chương 3: Kết quả nghiên cứu

Phần Kết luận là phần tổng kết luận án, bao gồm tóm lược mục đích nghiên cứu và các kết quả đạt được, đóng góp và hạn chế của luận án và các khuyến nghị.

Ngoài ba phần chính, luận án còn có danh sách các công trình khoa học của tác giả có liên quan đến luận án đã được công bố trên các tạp chí trong nước và kỷ yếu hội thảo quốc tế, danh mục tài liệu tham khảo và phụ lục.

CHƯƠNG 1: CƠ SỞ LÝ LUẬN VÀ TỔNG QUAN NGHIÊN CỨU

1.1. Cơ sở lý luận

1.1.1. Lý thuyết khảo thí hiện đại

Trước đây, lý thuyết khảo thí cổ điển (Classical Test Theory) đã ra đời từ khoảng cuối thế kỉ 19 và hoàn thiện vào khoảng những năm 1970, đã có nhiều đóng góp quan trọng, đặt nền móng cho hoạt động đánh giá trong giáo dục, nhưng cũng thể hiện một số điểm hạn chế như không thể tách biệt năng lực của thí sinh và các tham số của câu hỏi hay xem xét việc ứng đáp dựa vào cấp độ đề kiểm tra chứ không phải cấp độ câu hỏi (Lâm Quang Thiệp, 2010). Với mục đích khắc phục những hạn chế của khảo thí cổ điển, các nhà tâm trắc học đã cố gắng xây dựng một lý thuyết khảo thí hiện đại – lý thuyết ứng đáp câu hỏi, Item Response Theory – IRT, sử dụng mô hình toán học để dự đoán xác suất trả lời đúng một câu hỏi, dựa trên chỉ số về năng lực của người trả lời và độ khó của câu hỏi (Wu & Adams, 2007). Để đánh giá đối tượng nào đó thì lý thuyết khảo thí cổ điển tiếp cận ở cấp độ một đề kiểm tra, còn lý thuyết khảo thí hiện đại IRT tiếp cận ở cả cấp độ câu hỏi và đề thi. Lý thuyết khảo thí hiện đại đòi hỏi nhiều tính toán, nhưng nhờ sự tiến bộ vượt bậc của công nghệ tính toán bằng máy tính điện tử vào cuối thế kỉ 20 – đầu thế kỉ 21 nên nó đã phát triển nhanh chóng và đạt được những thành tựu quan trọng cho đến nay, được áp dụng rộng rãi, trong đó có phát triển đề thi, xây dựng ngân hàng câu hỏi, phân tích dữ liệu, trắc nghiệm thích ứng và so bằng đề thi (test equating) (Himelfarb, 2019).

1.1.1.1. Các giả thiết IRT

Szabo (2008) tổng lược ba giả thiết cơ bản làm nền tảng cho các mô hình IRT, bao gồm mối quan hệ giữa xác suất ứng đáp câu hỏi và năng lực thí sinh, tính đơn chiều và tính độc lập cục bộ.

Giả thiết đầu tiên liên quan đến mối quan hệ giữa các biến có thể quan sát được và biến tiềm ẩn, tức là mối quan hệ tương ứng giữa xác suất ứng đáp câu hỏi và năng lực của thí sinh. Mối quan hệ này được thể hiện bằng đường cong đặc trưng của câu hỏi (Item Characteristic Curve – ICC) (Baker, 1997, trích trong Szabo, 2008).

Giả thiết thứ hai là tính đơn chiều (unidimensionality). Keng (2008) đưa ra một ví dụ để minh họa, nếu IRT được sử dụng để mô hình hóa bài kiểm tra đọc hiểu thì người ta giả định rằng bất kỳ sự phụ thuộc thống kê nào giữa các câu trả lời đều được tính bằng năng lực đọc của thí sinh.

Giả thiết thứ ba của IRT là tính độc lập cục bộ (local independence) (Hambleton & Swaminathan, 1985, trích trong Keng, 2008). Tùy thuộc vào năng lực của thí sinh, xác suất ứng đáp với câu hỏi không phụ thuộc về mặt thống kê với xác suất ứng đáp với bất kỳ câu hỏi nào khác. Các học giả nhấn mạnh đặc tính quan trọng của IRT rằng nội dung của một câu hỏi không được cung cấp bất kỳ manh mối nào cho câu trả lời của một câu hỏi khác trong bài kiểm tra.

Việc hiểu và tuân thủ các giả thiết IRT là rất quan trọng vì chúng ảnh hưởng đến tính chính xác và độ tin cậy của quá trình xây dựng và phát triển các công cụ đo lường và đánh giá áp dụng IRT. Nếu có bất kỳ sự vi phạm nào đối với các giả thiết IRT, nhà nghiên cứu cần tiến hành sửa lỗi để cải thiện chất lượng của bài kiểm tra cũng như hiệu quả của quá trình đánh giá.

1.1.1.2. Các mô hình IRT

Một câu hỏi trắc nghiệm có 3 tham số đặc trưng. Đó là “độ khó” (kí hiệu là b), “độ phân biệt” (kí hiệu a) và “mức độ đoán mò” (kí hiệu c). Trong 3 tham số trên, tham số “độ khó (b)” là tham số quan trọng nhất của câu hỏi, tham số b sẽ được sử dụng để đối sánh với tham số năng lực (θ) của thí sinh. Tham số a được sử dụng để thể hiện đặc trưng phân biệt của câu hỏi và tham số c để chỉ tỉ lệ đoán mò của thí sinh khi gặp câu hỏi.

Hiện nay có ba mô hình phổ biến trong lý thuyết ứng đáp câu hỏi được phân loại theo số tham số đặc trưng mà mô hình xem xét, bao gồm mô hình một tham số kiểm tra các câu hỏi trắc nghiệm theo chỉ một tham số, độ khó của câu hỏi; mô hình hai tham số phân tích cả độ khó của câu hỏi và độ phân biệt câu hỏi, và mô hình ba tham số bao gồm độ khó của câu hỏi, độ phân biệt câu hỏi và mức độ dự đoán hay đoán mò câu trả lời. Cả ba mô hình đều sử dụng đường cong đặc trưng của câu hỏi làm căn cứ để phân tích. Bảng 1.1 trình bày một số thông tin cơ bản của ba mô hình IRT được tổng hợp (Carlson, 2020; Lâm Quang Thiệp, 2010). Trong đó, θ là mức năng lực của thí sinh trả lời câu hỏi và $P(\theta)$ là xác suất trả lời đúng câu hỏi của thí sinh có mức năng lực θ .

Các mô hình IRT một tham số, hai tham số và ba tham số cung cấp khả năng phân loại năng lực của thí sinh và hiểu rõ hơn về tính chất của các câu hỏi trong bài kiểm tra. Sự phức tạp của các mô hình tăng lên từ mô hình Rasch đến mô hình ba tham số, nhưng cũng cung cấp thông tin chi tiết và chính xác hơn về năng lực của thí sinh.

1.1.1.3. Các ứng dụng của IRT trong lĩnh vực kiểm tra đánh giá

Lý thuyết ứng đáp câu hỏi từ khi ra đời đã được ứng dụng rộng rãi vì tính hữu ích và ưu việt của nó; và việc phát minh ra máy tính cá nhân đã giúp nhiều nghiên cứu tiếp cận được sức mạnh tính toán cần thiết cho IRT.

Bảng 1.1: Các mô hình IRT (Carlson, 2020; Lâm Quang Thiệp, 2010)

Mô hình	Tham số	Biểu thức	Đường cong đặc trưng của câu hỏi
1 tham số	<ul style="list-style-type: none"> độ khó của câu hỏi b 	$P(\theta) = \frac{e^{\theta-b}}{1+e^{\theta-b}}$	<p>The graph shows three sigmoid curves for the Rasch (1PL) model. The x-axis is labeled 'Theta' and ranges from -3 to 3. The y-axis is labeled 'Probability' and ranges from 0.0 to 1.0. The legend indicates three curves for different difficulty levels: $b = -1.5$ (solid line), $b = 0.0$ (dashed line), and $b = 1.0$ (dash-dot line). The curves are shifted horizontally based on their difficulty level.</p>
2 tham số	<ul style="list-style-type: none"> độ khó của câu hỏi b độ phân biệt của câu hỏi a 	$P(\theta) = \frac{e^{a(\theta-b)}}{1+e^{a(\theta-b)}}$	<p>The graph shows three sigmoid curves for the 2PL model. The axes are the same as in the 1PL model. The legend indicates three curves for different combinations of discrimination (a) and difficulty (b): $a = 0.5, b = -1.5$ (solid line), $a = 1.8, b = 0.0$ (dashed line), and $a = 1.0, b = 1.0$ (dash-dot line). The curves vary in both their horizontal position and their slope.</p>
3 tham số	<ul style="list-style-type: none"> độ khó của câu hỏi b độ phân biệt của câu hỏi a mức độ đoán mò c 	$P(\theta) = c + (1-c) \frac{e^{a(\theta-b)}}{1+e^{a(\theta-b)}}$	<p>The graph shows three sigmoid curves for the 3PL model. The axes are the same as in the previous models. The legend indicates three curves for different combinations of discrimination (a), difficulty (b), and guessing (c): $a = 0.5, b = -1.5, c = .1$ (solid line), $a = 1.8, b = 0.0, c = .3$ (dashed line), and $a = 1.0, b = 1.0, c = .2$ (dash-dot line). The curves are vertically shifted due to the guessing parameter c.</p>

IRT được sử dụng để phát triển các bài kiểm tra chuẩn hóa, chẳng hạn như Bài kiểm tra năng lực học thuật (SAT). Sau đó, nó đã trở thành phương pháp tâm trắc quan trọng để xây dựng thang đo vì nó cung cấp một phương pháp giải quyết nhiều thách thức đo lường cần được giải quyết khi xây dựng một bài kiểm tra hoặc thang đo.

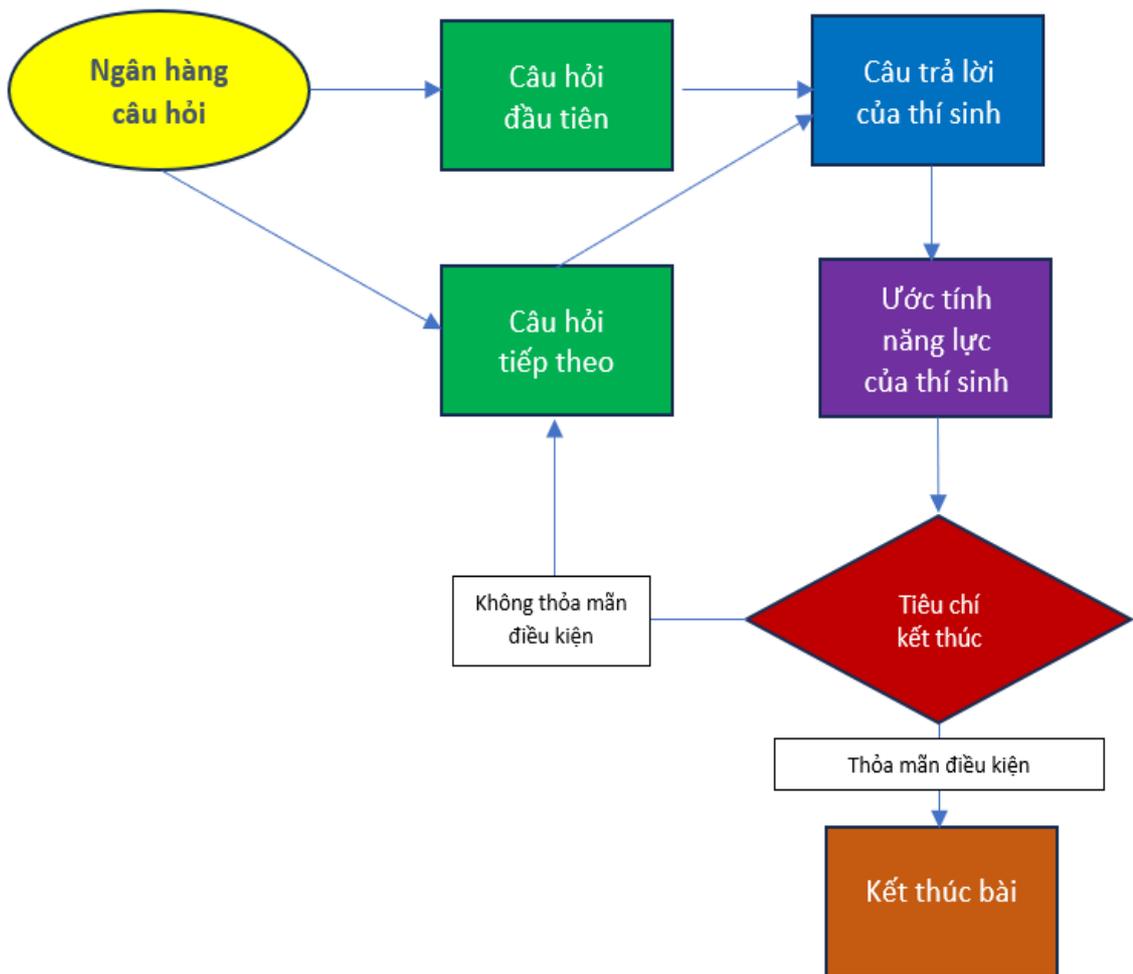
Mục đích của IRT là cung cấp một khuôn khổ để đánh giá mức độ thực hiện đánh giá và các câu hỏi riêng lẻ trong kiểm tra đánh giá. Một trong những ứng dụng phổ biến nhất của IRT là trong giáo dục, nơi các nhà nghiên cứu sử dụng IRT để phát triển và thiết kế các bài thi, xây dựng và duy trì ngân hàng câu hỏi và cân bằng độ khó của các câu hỏi trong bài thi cũng như các phiên thi khác nhau (Wu và cộng sự, 2016). Đầu tiên, IRT cung cấp những ước tính chính xác và đáng tin cậy hơn về khả năng của người làm bài kiểm tra, vì nó tính đến độ khó và sự phân biệt khác nhau của các câu hỏi cũng như điều chỉnh khả năng đoán mò và các yếu tố ngẫu nhiên khác. Thứ hai, IRT cho phép tạo ra các ngân hàng câu hỏi là tập hợp các câu hỏi được hiệu chuẩn trên thang đo chung nhờ phương pháp cân bằng, từ đó các câu kiểm tra có thể được chọn hoặc thiết lập theo các tiêu chí cụ thể. Thứ ba, IRT tạo điều kiện thuận lợi cho việc so sánh và cân bằng điểm số giữa các hình thức hoặc phiên bản khác nhau của bài kiểm tra, vì nó đặt chúng trên một thang điểm chung bất biến đối với các câu hỏi cụ thể được sử dụng. Thứ tư, IRT cho phép kiểm tra thích ứng, là một hình thức kiểm tra trên máy tính nhằm điều chỉnh độ khó và nội dung của các câu hỏi phù hợp với năng lực của người dự thi, dẫn đến các bài kiểm tra ngắn hơn và hiệu quả hơn.

Tổng thể, nhờ có một số lợi thế so với lý thuyết khảo thí cổ điển, IRT đã được ứng dụng rộng rãi trong việc phát triển và đánh giá bài kiểm tra. Ứng dụng của lý thuyết ứng đáp câu hỏi IRT không chỉ giúp nâng cao chất lượng của các kỳ thi khách quan hơn, chính xác hơn mà còn tối ưu hóa quá trình phát

triển, duy trì và triển khai. Điều này mang lại lợi ích lớn cho cả thí sinh và các bên liên quan.

1.1.2. Lý luận về trắc nghiệm thích ứng bằng máy tính

1.1.2.1. Trắc nghiệm thích ứng bằng máy tính



Hình 1.1: Biểu đồ CAT (Thompson & Weiss, 2011)

Hình 1.1 minh họa quy trình CAT – trắc nghiệm thích ứng bằng/trên máy tính (Thompson & Weiss, 2011). Trong một quy trình kiểm tra hoàn chỉnh, bài kiểm tra bắt đầu với một câu hỏi được chọn từ ngân hàng câu hỏi đã hiệu chuẩn. Câu hỏi đầu tiên này có thể được chọn ngẫu nhiên hoặc từ một nhóm câu hỏi có độ khó trung bình trong ngân hàng câu hỏi (Choi & McClenen, 2020; Oppl

và cộng sự, 2017). Nếu người dự thi đưa ra một câu trả lời đúng, thì một câu hỏi có độ khó cao hơn sẽ được chọn là câu hỏi tiếp theo, và ngược lại, khi người dự thi đưa ra một câu trả lời sai, một câu hỏi có độ khó thấp hơn sẽ được chọn là câu hỏi tiếp theo. Trong quá trình lặp lại này, khả năng của thí sinh được ước tính và tính toán lại dựa trên thành tích của thí sinh cho đến khi hệ thống thu thập đủ bằng chứng để xác định trình độ ngôn ngữ của thí sinh, nghĩa là đã thỏa mãn tiêu chí kết thúc.

1.1.2.2. Các thành tố của hệ thống trắc nghiệm thích ứng

Ngân hàng câu hỏi trắc nghiệm thích ứng

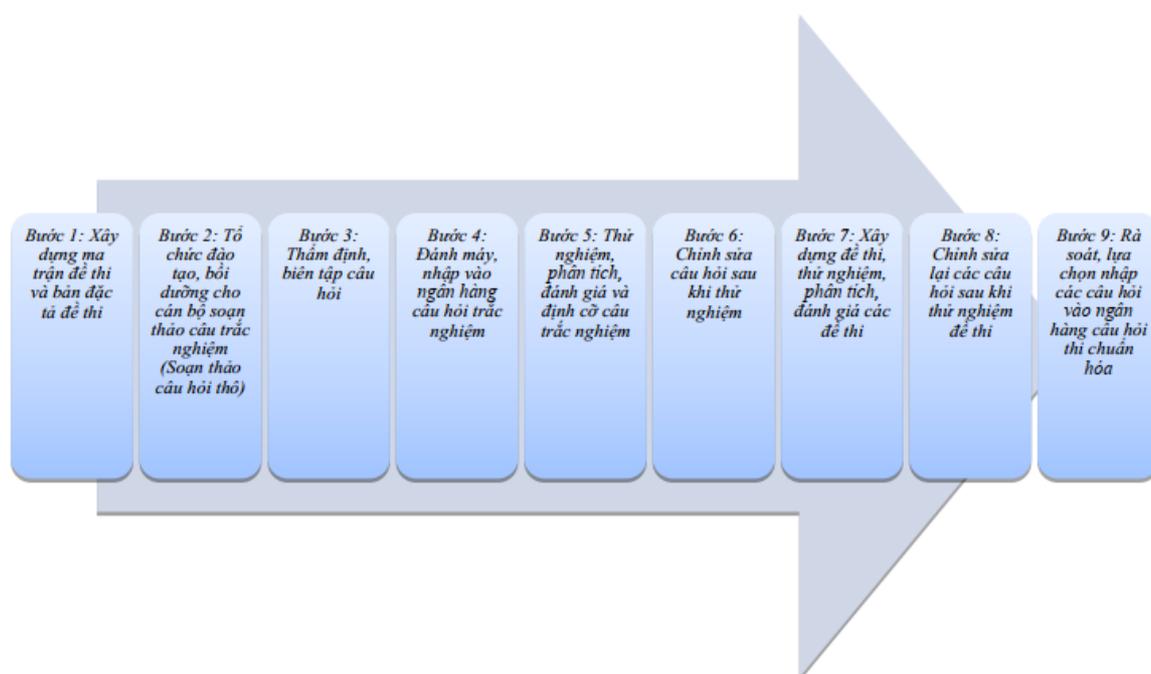
Thành tố đầu tiên của CAT là một ngân hàng câu hỏi đã được chuẩn hóa, dùng làm nội dung của hệ thống. Trong trường hợp đánh giá ngôn ngữ, ngân hàng câu hỏi bao gồm các câu hỏi về ngôn ngữ để tạo ra các bài kiểm tra ngôn ngữ. Với một hệ thống trắc nghiệm đã phát triển và đưa vào sử dụng thì các thuật toán đã được xác định từ trước, do đó, chất lượng của ngân hàng câu hỏi chất lượng đóng vai trò quyết định hiệu quả đánh giá năng lực của các thí sinh.

Tất cả các câu hỏi trong ngân hàng đều đầu tiên được phân tích với lý thuyết ứng đáp câu hỏi. Khi các câu hỏi đã được phân tích và xác định các tham số với lý thuyết ứng đáp, bộ câu hỏi sẽ được hiệu chuẩn với phương pháp cân bằng, nghĩa là các tham số của câu hỏi phải được đưa về cùng một thang đo, sau đó ngân hàng câu hỏi sẽ được biên tập và lưu trữ kèm theo các tham số thống kê của chúng, sẵn sàng cho việc thực hiện các thuật toán sau này trong hệ thống (Choi & McClenen, 2020).

Thompson và Weiss (2011) nhấn mạnh sự cần thiết của việc xây dựng ngân hàng câu hỏi không chỉ cần lưu ý đến số lượng câu hỏi trong ngân hàng, mà còn đến sự phân bố của các thông số câu hỏi và những cân nhắc thực tế như

phân phối nội dung và các dự đoán về mức độ phân phối từng câu hỏi. Các tác giả cũng cho rằng việc xây dựng ngân hàng câu hỏi cần dựa trên những nghiên cứu thực nghiệm, cụ thể là tiến hành thử nghiệm bộ câu hỏi. Nhờ đó, các tham số của câu hỏi được ước tính thông qua phân tích thống kê về phản hồi thực tế của thí sinh đối với câu hỏi.

Ở Việt Nam, số lượng công bố xây dựng ngân hàng câu hỏi còn hạn chế. Trong đó có thể kể đến các nghiên cứu của gần đây như Le và cộng sự (2019), Le và Nguyen (2021), Nguyen và cộng sự (2021), Nguyen và Nguyen (2020). Các nghiên cứu có lưu ý đến một quy trình nghiêm túc để xây dựng ngân hàng câu hỏi như Hình 1.2, việc áp dụng mô hình IRT để phát triển ngân hàng câu hỏi, tuy nhiên các nghiên cứu này đều chưa có những báo cáo cụ thể liên quan đến quá trình cân bằng đề thi thử nghiệm để đảm bảo các tham số của các câu hỏi trong ngân hàng đã được đưa về cùng một thang đo.



Hình 1.2: Quy trình xây dựng ngân hàng câu hỏi
(Lê Thái Hưng và cộng sự, 2019)

Các thuật toán trắc nghiệm thích ứng

Các thành phần khác của CAT là các thuật toán CAT quyết định câu hỏi đầu tiên (điểm khởi đầu), chọn câu hỏi tiếp theo (thuật toán lựa chọn câu hỏi), tính điểm các câu trả lời đúng để dự đoán năng lực của thí sinh (thuật toán tính điểm), và kiểm tra tiêu chí đã định trước để kết thúc bài kiểm tra (tiêu chí kết thúc) (Thompson & Weiss, 2011).

- Điểm khởi đầu

Có một số tùy chọn có sẵn như ước tính năng lực ban đầu θ được chỉ định cho mỗi thí sinh trước khi một câu hỏi được đưa ra. Đơn giản nhất là chỉ định một giá trị cố định tương ứng với điểm trung bình. Với IRT, mức này thường là 0,0.

- Thuật toán lựa chọn câu hỏi

Thuật toán lựa chọn câu hỏi rất quan trọng vì nó không chỉ đề cập đến các tính toán cụ thể để xác định câu hỏi thích hợp nhất mà còn liên quan đến tác động của các ràng buộc thực tế. Lựa chọn câu hỏi thường dựa trên thông tin câu hỏi, nhằm tìm cách định lượng nhận định một số câu hỏi phù hợp hơn những câu hỏi khác trong một tình huống nhất định. Ví dụ, sẽ không có ý nghĩa gì khi giao một câu hỏi rất dễ cho một thí sinh khá; thí sinh gần như có thể đảm bảo sẽ trả lời chính xác. Kết quả tương tự với trường hợp câu hỏi quá khó với những người có năng lực thấp.

- Thuật toán tính điểm (ước tính năng lực)

Hầu hết các CAT sử dụng IRT để chấm điểm, ngoài việc lựa chọn câu hỏi ở bước trên.

- Tiêu chí kết thúc

Các bài kiểm tra thích ứng trên máy tính có thể được thiết kế với số lượng câu hỏi cố định hoặc thay đổi. Một bài kiểm tra với CAT không chỉ thích ứng độ khó câu hỏi cho thí sinh, mà còn thích ứng với số lượng các câu hỏi cần thiết. Có nhiều phương pháp khác nhau để thực hiện điều này. Một số xem xét ước tính năng lực của thí sinh, một số khác xem xét sai số chuẩn của phép ước lượng năng lực và độ lớn của ngân hàng câu hỏi.

1.1.2.3. Khung xây dựng đề kiểm tra thích ứng

Khung xây dựng của Thompson và Weiss (2011) được đưa ra dựa trên việc đối chiếu các phương pháp nghiên cứu hiện hành từ một lượng lớn nghiên cứu về các khía cạnh kỹ thuật của kiểm tra thích ứng trên máy tính trong suốt 40 năm để cung cấp những chỉ dẫn hữu dụng với việc xây dựng bất cứ đề kiểm tra thích ứng nào. Khung được chia làm năm bước như Bảng 1.2.

Bảng 1.2: Khung xây dựng CAT (Thompson & Weiss, 2011)

Bước	Giai đoạn	Nhiệm vụ chính
1	Nghiên cứu tính khả thi, khả năng ứng dụng và lập kế hoạch	Mô phỏng Monte Carlo, đánh giá trường hợp thương mại
2	Xây dựng nội dung ngân hàng câu hỏi hoặc sử dụng ngân hàng câu hỏi sẵn có	Thiết kế và chỉnh sửa câu hỏi
3	Thử nghiệm và định cỡ câu hỏi	Thử nghiệm, phân tích câu hỏi
4	Xác định thông số kỹ thuật cho CAT	Mô phỏng phân tích sau hoặc song song
5	Công bố CAT	Công bố và phổ biến, phát triển phần mềm

Bước 1: Nghiên cứu tính khả thi, khả năng ứng dụng và lập kế hoạch

Giai đoạn đầu tiên trong quá trình phát triển CAT là xác định xem phương pháp CAT có khả thi hay không. CAT là một lựa chọn rất hấp dẫn về với một số lợi thế rõ rệt nhưng việc chuyển đổi từ kiểm tra truyền thống sang CAT có thể là khá mạo hiểm, không chỉ từ góc độ phản ứng tâm lý mà còn từ góc độ quản lý nguồn lực. Thompson và Weiss (2011) nhận định quyết định chuyển đổi hình thức đánh giá từ các bài kiểm tra dạng cố định sang CAT không phải là một quyết định dễ dàng.

Các cân nhắc được liệt kê bao gồm tổ chức có chuyên môn về đo lường đánh giá không, hay có đủ khả năng chi trả nếu sử dụng chuyên gia tư vấn bên ngoài không; tổ chức có đủ năng lực để phát triển các ngân hàng câu hỏi lớn không; công cụ tiến hành CAT có sẵn để sử dụng hay tổ chức có đủ nguồn lực để phát triển công cụ của riêng mình không; việc chuyển đổi bài kiểm tra sang CAT có khả năng làm giảm độ dài bài kiểm tra dự kiến không; việc giảm độ dài bài kiểm tra có chuyển thành tiết kiệm thời gian làm bài của thí sinh, để chuyển thành tiết kiệm chi phí không; hoặc ngay cả khi CAT có chi phí cao hơn và không làm giảm đáng kể thời gian làm bài của thí sinh, thì CAT có gia tăng độ chính xác và bảo mật để bù lại không.

Bước 2: Xây dựng nội dung ngân hàng câu hỏi

Khi quyết định cuối cùng đã được đưa ra để chuyển đổi sang CAT, bước tiếp theo là thành lập một ngân hàng câu hỏi. Thompson và Weiss (2011) nhấn mạnh một lần nữa sự cần thiết của việc thực hiện bước 2 dựa trên những nghiên cứu thực nghiệm.

Bước này cần lưu ý không chỉ đến số lượng câu hỏi trong ngân hàng, mà còn đến sự phân bố của các thông số câu hỏi và những cân nhắc thực tế như phân phối nội dung và các dự đoán về độ phân biệt của từng câu hỏi.

Bất kể ngân hàng sẽ bao gồm tất cả các câu hỏi mới hay kết hợp giữa cũ và mới, điều quan trọng là phải xem xét các số liệu thống kê của các câu hỏi trong một đề kiểm tra. Bước này cũng cần chú ý đến mục tiêu đề ra của đề kiểm tra. Mục tiêu cao có thể dẫn tới quá trình phát triển ngân hàng câu hỏi loại bỏ một tỷ lệ phần trăm đáng kể các câu hỏi, từ đó yêu cầu tăng số lượng câu hỏi thô để đảm bảo số lượng câu hỏi đạt yêu cầu sau quá trình hiệu chỉnh.

Bước 3: Thử nghiệm, hiệu chỉnh và cân bằng

Sau khi các câu hỏi được thiết kế, bước tiếp theo là tiến hành thử nghiệm. Bước này rất quan trọng và cần thiết đối với CAT vì các câu hỏi cần được đối sánh để kiểm tra dựa trên các thông số của IRT và các thông số được ước tính thông qua phân tích thống kê về phản hồi thực tế của thí sinh đối với câu hỏi. Kích thước mẫu cần thiết cho việc kiểm tra thử tùy thuộc vào mô hình IRT được sử dụng. Yoes (1995, được trích dẫn trong Thompson & Weiss, 2011) gợi ý rằng cần 500 đến 1000 thí sinh cho mỗi câu hỏi đối với mô hình IRT ba tham số.

Sau khi tiến hành thử nghiệm, các thông số của câu hỏi phải được ước tính bằng phần mềm hiệu chuẩn dựa trên IRT. Một nhiệm vụ quan trọng của bước này là cân bằng, đảm bảo rằng các thông số của tất cả các câu hỏi được hiệu chuẩn trên cùng một thước đo.

Bước 4: Xác định thông số kỹ thuật cho CAT

Tại thời điểm này, một ngân hàng câu hỏi đã được phát triển và hiệu chỉnh với IRT. Tuy nhiên, đây chỉ là thành tố đầu tiên trong số năm thành tố

cấu tạo của CAT được mô tả trước đây. Trước khi CAT có thể được công bố và đưa vào sử dụng, bốn thành tố còn lại sau đây phải được xác định.

- Điểm khởi đầu
- Thuật toán lựa chọn câu hỏi
- Thuật toán tính điểm (ước tính năng lực)
- Tiêu chí kết thúc

Bước 5: Công bố CAT

Khi các thông số kỹ thuật cho tất cả các thành phần cần thiết đã được xác định, cũng như bất kỳ thuật toán bổ sung nào, CAT cuối cùng có thể được công bố. Nếu phần mềm phân phối và phát triển CAT đã tồn tại (được mua, hoặc cấp quyền truy cập), thì bước này sẽ ít khó khăn. Tuy nhiên, nếu tổ chức đang phát triển nền tảng của riêng mình, thì bước này có thể là khá khó khăn. Tuy nhiên, nếu trường hợp đó xảy ra, hầu hết nhiệm vụ phát triển có thể được thực hiện đồng thời với bốn bước trước đó, tiết kiệm một lượng thời gian đáng kể.

Ngoài năm bước được nêu ở trên, Thompson và Weiss (2011) cũng bổ sung thêm một nhiệm vụ quan trọng là bảo trì CAT, bao gồm (1.) việc kiểm tra là liệu kết quả CAT thực tế sau khi công bố có khớp với kết quả mong đợi hay không; và (2.) việc “làm mới” ngân hàng câu hỏi thi bằng cách đưa các câu hỏi mới vào ngân hàng câu hỏi.

1.1.3. Lý luận về đánh giá kiến thức từ vựng tiếp nhận tiếng Anh

1.1.3.1. Từ vựng và kiểm tra từ vựng

Với những người ngoài lĩnh vực nghiên cứu ngôn ngữ, việc đánh giá xem một ai đó có biết hoặc không biết một từ hay nhiều từ trong ngôn ngữ khác

là không hề khó khăn (Read, 2019). Cách đơn giản được nghĩ đến là đưa cho người đó một từ trong ngôn ngữ nguồn và yêu cầu từ tương đương trong ngôn ngữ khác hay còn gọi là ngôn ngữ đích. Nếu người này đưa ra một từ chính xác, có thể đánh giá người đó biết ngôn ngữ đích. Tuy nhiên, trên thực tế việc kiểm tra đánh giá từ vựng không hề đơn giản mà đòi hỏi việc xem xét nhiều khía cạnh của từ vựng (Schmitt, 2014).

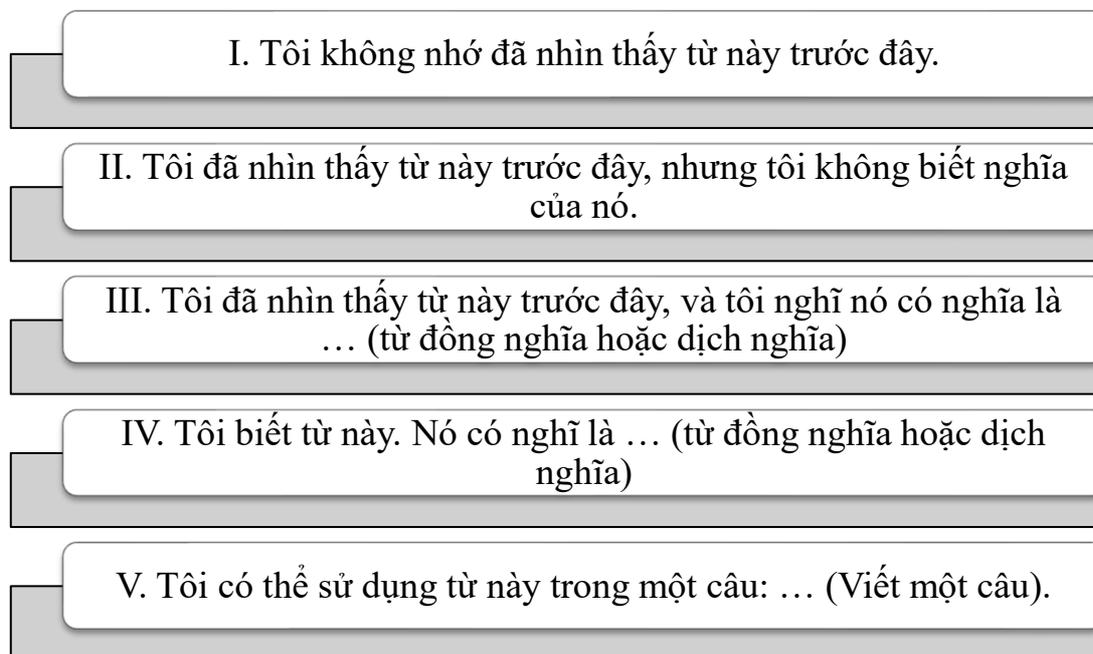
Khi bàn về từ vựng, các nhà nghiên cứu đưa ra các cặp khái niệm có tính chất phân biệt bao gồm: (1) từ vựng tiếp nhận (receptive vocabulary) và từ vựng sản sinh (productive vocabulary); (2) kiến thức thụ động (passive knowledge) và kiến thức chủ động (active knowledge), (3) số lượng từ vựng (breadth) và chiều sâu hiểu biết về từ vựng (depth).

Nation (2013) đưa ra phân biệt giữa từ vựng tiếp nhận (receptive vocabulary) và từ vựng sản sinh (productive vocabulary). Cặp khía cạnh này hướng tới mối liên hệ giữa từ vựng với các nhóm kỹ năng sử dụng tiếng Anh. Từ vựng tiếp nhận liên quan đến việc nhận thức được hình thức từ trong khi nghe và đọc và thu nhận được nghĩa của từ đó; từ vựng sản sinh liên quan đến việc mong muốn diễn đạt được ý nghĩa thông qua nói và viết, thu nhận và sản sinh được hình thức phù hợp của từ ở dạng nói hoặc viết.

Laufer và cộng sự (2004) định nghĩa kiến thức chủ động liên quan đến việc có thể đưa ra đúng dạng từ (form) (thể nói hoặc viết) trong ngôn ngữ đích trong khi kiến thức thụ động liên quan đến việc biết nghĩa (meaning) của một từ ở ngôn ngữ đích. Ví dụ: nếu một người nói tiếng Anh không phải là bản ngữ được yêu cầu nói hoặc viết một từ tiếng Anh có nghĩa “when solid becomes liquid” và đưa ra câu trả lời là “melt”, người đó đã thể hiện được kiến thức chủ động. Nếu người đó được đưa cho từ “melt” và giải thích nghĩa của từ đó là “when something turns into water/liquid” thì người đó đã chứng thực được kiến

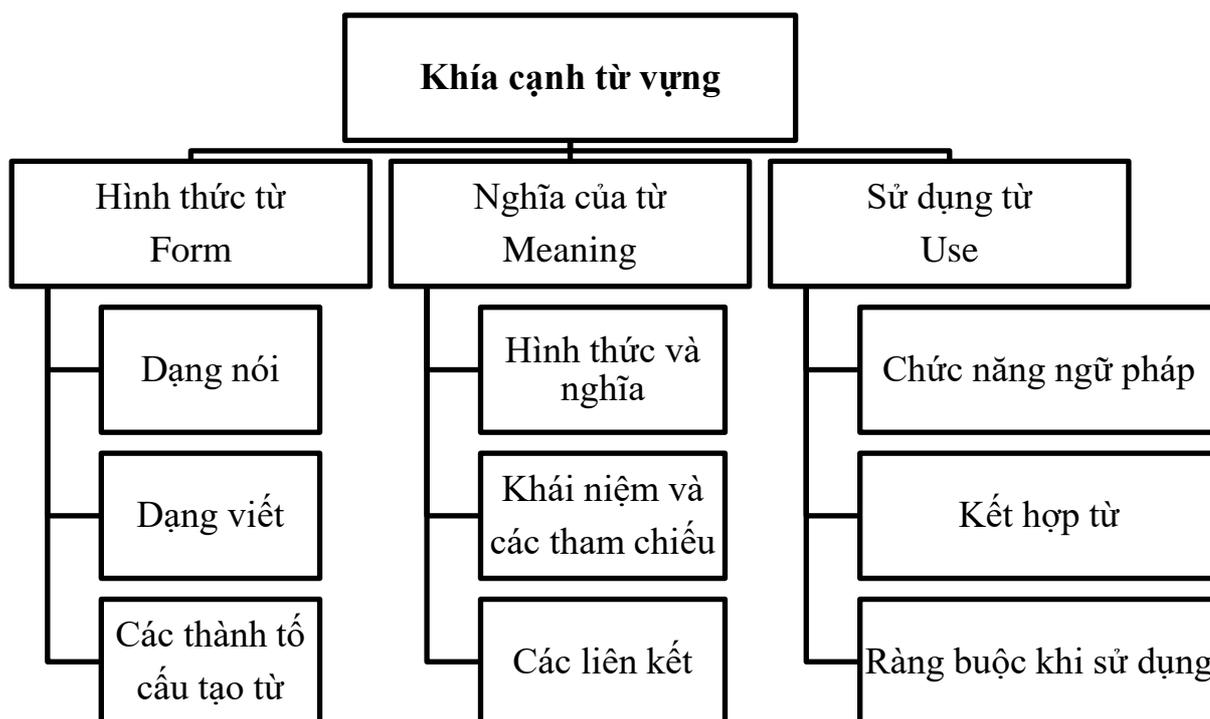
thức thụ động của mình về từ “melt”. Cặp khái niệm này có liên hệ chặt chẽ đến một cặp khái niệm khác, đó là khả năng nhận biết (recognition) và khả năng hồi suy (recall). Theo Laufer và cộng sự (2004) khả năng nhận biết được thể hiện thông qua việc nhận diện được từ trong một số các lựa chọn được cung cấp, còn khả năng hồi suy được thể hiện thông qua việc tự đưa được từ mà không có các phương án cho trước. Ví dụ: một người được đưa ra danh sách các từ khác nhau và được hỏi từ nào có nghĩa “when solid becomes liquid”, người đó chọn được từ “melt” sẽ thể hiện được khả năng nhận biết từ “melt”; nếu danh sách các từ không được cung cấp mà người đó vẫn đưa ra câu trả lời “melt”, khả năng hồi suy của người đó sẽ được ghi nhận.

Cặp khái niệm được cho là phổ biến khác trong nghiên cứu từ vựng, đặc biệt trong kiểm tra từ vựng là lượng từ - size (hoặc chiều rộng - breadth) và mức độ hiểu biết - strength (hoặc chiều sâu - depth) (Read, 2019; Schmitt, 2014). Lượng từ hay độ rộng từ vựng đề cập đến số lượng từ mà một người biết và đã được chứng minh là một minh chứng có giá trị về khả năng ngôn ngữ tổng thể (Milton, 2009). Mặt khác, sự hiểu biết / độ sâu từ vựng đề cập đến mức độ hiểu biết của một từ (hoặc một nhóm từ). Trong tài liệu ngôn ngữ học ứng dụng, thuật ngữ chiều sâu của kiến thức từ vựng có những cách hiểu khác nhau. Một số tác giả (Paribakht & Weshe, 1997; Schmitt & Zimmerman, 2002) đã đưa ra một cách tiếp cận phát triển đối với chiều sâu của kiến thức từ. Đối với các tác giả này, kiến thức từ chuyên sâu bao gồm từ việc nhận biết đơn thuần với từ đã gặp trước đó, đến việc có thể sử dụng từ một cách hiệu quả và theo cách phù hợp với ngữ cảnh. Một trong những cách đánh giá nổi tiếng nhất về độ sâu của kiến thức từ trong cách tiếp cận phát triển này là Thang kiến thức từ vựng (Vocabulary Knowledge Scale) của Paribakht và Welshe (1997), được trình bày trong Hình 1.3.



Hình 1.3: Thang đo kiến thức từ vựng (Paribakht & Welshe, 1997)

Các nhà ngôn ngữ học khác (Nation, 2013; Schmitt, 2010) đã khái niệm hóa chiều sâu của kiến thức từ vựng không phải theo hướng tiếp cận phát triển, mà thay vào đó là hướng tiếp cận theo khía cạnh, cụ thể là người học có thể được đánh giá là biết một từ ở mức độ thành thạo một số khía cạnh khác nhau liên quan đến nó. Nation (2013) đã xác định một khung đánh giá từ vựng (Hình 1.4) bao gồm ba khía cạnh chính, mỗi khía cạnh bao gồm ba khía cạnh nhỏ: (1) hình thức từ (các khía cạnh nhỏ: dạng nói, dạng viết và các thành tố của từ), (2) nghĩa của từ (các khía cạnh nhỏ: hình thức và ý nghĩa, khái niệm và các tham chiếu, và các liên kết), và (3) sử dụng từ (các khía cạnh nhỏ: chức năng ngữ pháp, kết hợp từ và các ràng buộc khi sử dụng).



Hình 1.4: Các khía cạnh của kiến thức từ vựng (Nation, 2013)

Các nhà nghiên cứu nhận định việc đánh giá đồng thời độ rộng hay lượng từ và độ sâu hay mức độ hiểu biết từ vựng là một nhiệm vụ khó đạt được. Tùy thuộc các mục đích đào tạo, đánh giá hay nghiên cứu khác nhau, trọng tâm có thể hướng tới lượng từ hay mức độ hiểu biết từ vựng, từ đó đánh giá đầy đủ và chính xác năng lực về từ vựng của đối tượng đang hướng tới.

1.1.3.2. Đánh giá kiến thức từ vựng tiếp nhận tiếng Anh

Từ vựng tiếp nhận tiếng Anh được xem là khía cạnh cơ bản nhất và quan trọng nhất của kiến thức từ vựng trong kiểm tra đánh giá, đó là mối quan hệ giữa dạng từ (form) và nghĩa (meaning), khía cạnh này làm nền móng để tiến hành việc học tập và lĩnh hội các khía cạnh khác của từ vựng (Webb & Chang, 2012).

Trong luận án, khái niệm kiến thức từ vựng tiếp nhận tiếng Anh được sử dụng mang tính đại diện cho những đặc điểm cụ thể hơn của từ vựng đã được tóm lược trong phần trên, cụ thể: (1) kiến thức từ vựng tiếp nhận tiếng Anh hướng tới khía cạnh dạng từ và nghĩa trong các kỹ năng tiếp nhận là đọc và nghe, do đó kiến thức từ vựng tiếp nhận tiếng Anh thực chất là kiến thức thụ động của người được kiểm tra, tập trung vào kỹ năng nhận biết nghĩa của một từ cho trước ở dạng viết (written form) hoặc dạng nói (spoken form); (2) kiến thức từ vựng tiếp nhận tiếng Anh hướng tới một mức độ duy nhất trong chiều sâu của kiến thức từ vựng, đó là mức độ biết nghĩa, mức độ giữa trong thang đo kiến thức từ vựng theo cách tiếp cận phát triển của Paribakht & Welshe (1997); do đó, việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh thực chất là xác định độ rộng của từ vựng, có thể được thực hiện với một danh sách từ được lựa chọn phù hợp với đối tượng kiểm tra để xác định số lượng từ người được kiểm tra đã nhận diện được nghĩa một cách chính xác, từ đó có thể phục vụ các mục đích khác nhau của kiểm tra từ vựng mà Nation (2013) liệt kê:

1. để tìm ra những vấn đề người học đang gặp khó khăn để định hướng giải pháp (kiểm tra chẩn đoán). Mục đích này cũng có thể liên quan đến việc xem xét người học có thể sử dụng các chiến lược học từ vựng và xử lý tình huống như thế nào;
2. xếp người học vào các lớp đúng trình độ (kiểm tra xếp lớp);
3. để xem liệu một nhóm từ đã học gần đây đã được học chưa (các bài kiểm tra kết quả ngắn hạn hay kiểm tra quá trình);
4. để xem liệu một khóa học có thành công trong việc dạy các từ cụ thể hay không (các bài kiểm tra kết quả dài hạn hay kiểm tra tổng kết); và
5. để xem kiến thức từ vựng của người học thế nào (đánh giá trình độ người học). (tr. 515)

Trước hết và quan trọng nhất, việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh có mối liên quan rõ rệt đến việc xác định trình độ thông thạo ngôn ngữ của người được kiểm tra vì kiến thức từ vựng là “nền tảng cho tất cả việc sử dụng ngôn ngữ” (Schmitt và cộng sự, 2017) và do đó là một phần quan trọng của việc học ngôn ngữ. Đạt được một mức trình độ và hiểu biết nhất định của kiến thức từ vựng là một trong những điều kiện tiên quyết quan trọng để học ngôn ngữ thành công. Điều này làm cho kiến thức từ vựng trở thành một yếu tố dự báo mạnh mẽ về trình độ ngôn ngữ của người học và thậm chí cả thành tích học tập của họ (Lin & Morrison, 2010). Có được kết quả đánh giá kiến thức từ vựng tiếp nhận tiếng Anh, người học có thể tự xác định được trình độ của mình, làm nền tảng để tìm ra những khó khăn thiếu sót để hỗ trợ và nâng chất lượng của quá trình học tập. Trong khi đó, năng lực từ vựng của người học có xu hướng cải thiện khi trình độ ngôn ngữ của họ phát triển (Zareva và cộng sự, 2005) kể từ quá trình áp dụng bốn kỹ năng ngôn ngữ là đọc, nghe, nói và viết trong giao tiếp luôn có lợi cho việc thu nhận các từ mới học vào bộ nhớ (Laufer và cộng sự, 2004). Qian & Lin (2019), sau khi phân tích mối liên hệ giữa kiến thức từ vựng với cả bốn kỹ năng nghe, nói, đọc, viết, đã đề xuất các nghiên cứu tiếp theo về mối liên hệ giữa từ vựng, bao gồm cả độ rộng và độ sâu của từ vựng, với trình độ thành thạo ngôn ngữ. Kết quả đánh giá kiến thức từ vựng tiếp nhận tiếng Anh cũng có thể sử dụng cho các mục đích khác như xếp lớp cũng như đánh giá hiệu quả của một bài học hay một khóa học.

Xét một cách tổng thể, trong luận án, đánh giá kiến thức từ vựng tiếp nhận tiếng Anh được thực hiện với một danh sách từ vựng phù hợp được lựa chọn để xác định số lượng từ mà người được kiểm tra có thể nhận diện được nghĩa một cách chính xác. Việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh đóng một vai trò lớn trong cả đào tạo và nghiên cứu. Trong đào tạo, các bài kiểm tra về kiến thức từ vựng nói chung, bao gồm từ vựng tiếp nhận tiếng Anh

có thể thúc đẩy việc học tập, nâng cao nhận thức về các khía cạnh khác nhau của từ vựng (Yanagisawa & Webb, 2019), chỉ ra mức độ phát triển từ vựng trong một khóa học (Kremmel, 2019) và tiết lộ những từ học sinh biết và những từ nào học sinh cần học (Gyllstad, 2019). Trong nghiên cứu, chất lượng và trọng tâm của các bài kiểm tra kiến thức từ vựng có thể ảnh hưởng đến việc học từ vựng có diễn ra hay không, cũng như số lượng từ mà người tham gia được đánh giá là “biết” (Read, 2019).

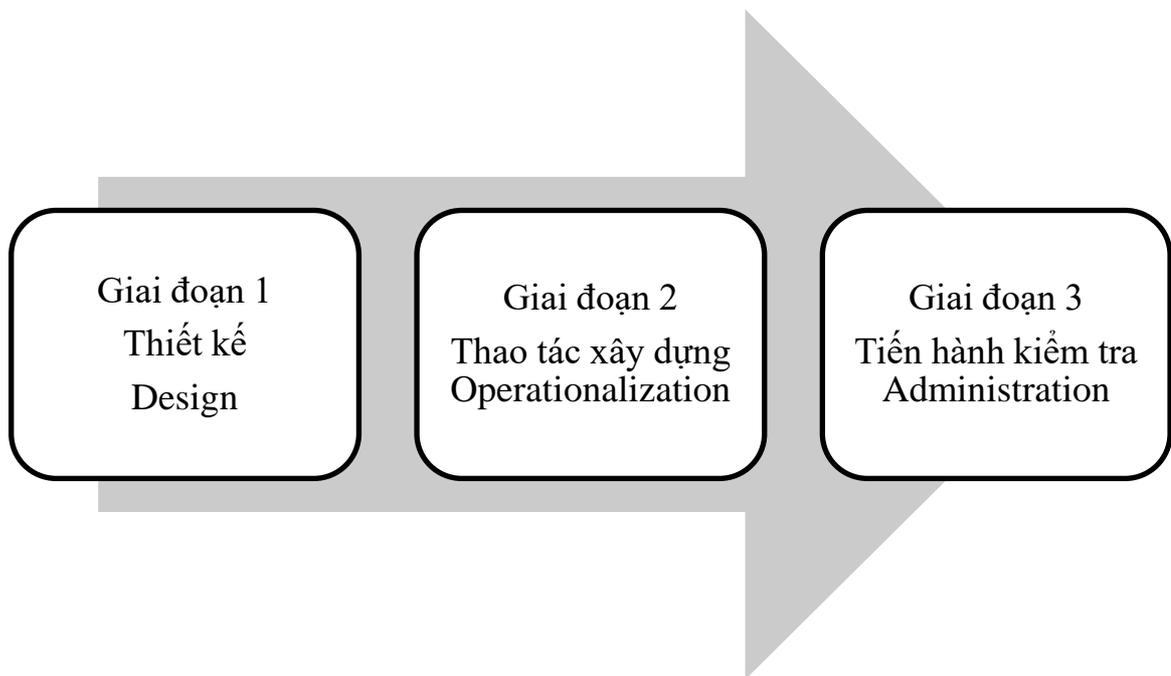
1.1.3.3. Khung xây dựng đề kiểm tra ngôn ngữ

Khung phát triển đề kiểm tra ngôn ngữ được xem là nền tảng của rất nhiều các nghiên cứu trong nhiều thập kỷ qua là của Bachman và Palmer (1996) với ba giai đoạn - thiết kế, thao tác xây dựng và tiến hành kiểm tra.

Giai đoạn 1: Thiết kế đề kiểm tra

Giai đoạn đầu tiên của quá trình phát triển đề kiểm tra hướng tới một “bản trình bày thiết kế” bao gồm một loạt các đầu mục:

1. mô tả mục đích của đề kiểm tra,
2. mô tả các loại hình lĩnh vực và nhiệm vụ về sử dụng ngôn ngữ đích,
3. mô tả đối tượng thí sinh mà bài kiểm tra hướng tới,
4. định nghĩa (các) trọng điểm được kiểm tra,
5. kế hoạch đánh giá các tiêu chí tính hữu dụng của đề thi, và
6. bản kê các nguồn lực và điều kiện sẵn có cùng với kế hoạch phân phối và quản lý nguồn lực



Hình 1.5: Quy trình xây dựng đề kiểm tra (Bachman & Palmer, 1996)

Giai đoạn 2: Thao tác xây dựng

Giai đoạn thao tác xây dựng hướng tới việc thiết lập được một “bản đặc tả thiết kế” gồm các loại hình nhiệm vụ kiểm tra và một bản thiết kế cấu trúc đề thi miêu tả cách thức tổ chức các nhiệm vụ kiểm tra để tạo một đề kiểm tra thực sự. Cấu trúc đề thi có các nội dung sau: số lượng phần/nhiệm vụ; sự tách biệt giữa các phần, trình tự các phần, mức quan trọng tương quan giữa các phần, số nhiệm vụ mỗi phần. Đặc tả thiết kế nhiệm vụ có các nội dung sau: mục đích, trọng điểm nội dung/kỹ năng, định dạng, phân định thời gian, chỉ dẫn làm bài, đặc tính của câu hỏi và câu trả lời, cách thức tính điểm.

Giai đoạn 3: Tiến hành kiểm tra

Giai đoạn này bao gồm việc thực hiện kiểm tra với một nhóm thí sinh, thu thập thông tin và phân tích thông tin này với hai mục tiêu:

- Đánh giá tính hữu dụng của bài kiểm tra, và
- Đưa ra những suy luận và quyết định mà bài kiểm tra hướng tới

Giai đoạn này thường được thực hiện với hai bước: thử nghiệm (try-out) và thực thi (operational testing). Thử nghiệm là tiến hành kiểm tra với mục đích thu thập thông tin về tính hữu dụng của đề kiểm tra và để chỉnh sửa đề kiểm tra cũng như quy trình kiểm tra. Thực thi là tiến hành kiểm tra thực sự để đạt được mục đích của bài kiểm tra đề ra, đồng thời thu thập thông tin về tính hữu ích của bài kiểm tra.

Giai đoạn này cần chú ý đến quy trình tiến hành kiểm tra và thu thập phản hồi, quy trình phân tích điểm số và quá trình lưu trữ để tạo lập ngân hàng thi.

Cần phải lưu ý rằng Bachman và Palmer (1996) có đưa ra một khái niệm “tính hữu dụng” (test usefulness) của đề thi để chi phối toàn bộ quá trình xây dựng đề thi. Khái niệm này được định nghĩa bao gồm sáu nội dung: (1.) tính thực tiễn (practicality), (2.) độ tin cậy (reliability), (3.) tính xác thực (authenticity), (4.) tính tương tác (interactiveness), (5.) độ giá trị cấu trúc (construct validity) và (6.) tác động (impact). Các nội dung này cần được xem xét chi tiết trong quá trình xác trị đề kiểm tra.

Khung xây dựng đề kiểm tra ngôn ngữ của Bachman và Palmer có giá trị vô cùng to lớn, đặt nền móng cho các hoạt động cũng như nghiên cứu về kiểm tra đánh giá ngôn ngữ trong nhiều năm qua, và luôn được các học giả đánh giá cao về giá trị nghiên cứu cũng như giá trị thực tiễn.

1.1.3.4. Khung kiểm tra từ vựng tiếng Anh

Phần này trình bày lại khung kiểm tra từ vựng của Read và Chapelle (2001) và tóm lược lại phân giải thích năm thành tố của khung, cụ thể là mục đích kiểm tra (test purpose), các cân nhắc độ giá trị (validity considerations),

các yếu tố trung gian (mediating factors), thiết kế đề kiểm tra (test design), và xác trị (validation).

Mục đích kiểm tra và những cân nhắc về độ giá trị

Mục đích kiểm tra được xác định gồm ba yếu tố thuộc hàng đầu tiên trong khung kiểm tra từ vựng. Ba yếu tố này được phân loại cụ thể hơn trong Bảng 1.3.

Suy luận và độ giá trị cấu trúc

Các suy luận đề cập đến các kết luận rút ra về khả năng ngôn ngữ dựa trên mức độ thể hiện của thí sinh trong bài kiểm tra. Rõ ràng, mức độ kiến thức từ vựng của người học không thể quan sát trực tiếp mà được suy ra từ tỷ lệ câu trả lời đúng cho các câu hỏi trong mỗi phần của bài kiểm tra. Các suy diễn có thể ở nhiều cấp độ khác nhau: (1) cấp độ câu hỏi, (2) cấp độ theo một phần đề và (3) cấp độ toàn đề kiểm tra.

Bảng 1.3: Các yếu tố xác định mục đích kiểm tra

Các yếu tố của mục đích kiểm tra	
1) Suy luận rút ra từ kết quả kiểm tra	cấp độ câu hỏi
	cấp độ theo phần đề kiểm tra
	cấp độ toàn đề kiểm tra
2) Việc sử dụng kết quả kiểm tra	dạy và học: xếp lớp, đánh giá tiến bộ, chẩn đoán, đánh giá trình độ
	nghiên cứu
	đánh giá
3) Các tác động bài kiểm tra dự kiến mang lại	

Công dụng: mức độ liên quan và tính hữu dụng

Yếu tố thứ hai của mục đích kiểm tra, công dụng, đề cập đến việc sử dụng kết quả kiểm tra, nói cách khác là trả lời câu hỏi kết quả của bài kiểm tra được sử dụng để làm gì. Read và Chapelle (2001) chia thành ba công dụng theo Bachman và Palmer (1996) nhưng có sử dụng thuật ngữ khác đi. Ba công dụng đó được liệt kê là (1.) Công dụng trong giảng dạy liên quan đến việc ra quyết định về người học, phổ biến nhất là trong bối cảnh giáo dục. Đây là công dụng được biết đến rộng rãi nhất của các bài kiểm tra ngôn ngữ để xếp lớp, đánh giá tiến bộ người học hay đánh giá trình độ thông thạo của người học; (2.) Công dụng trong nghiên cứu liên quan đến vai trò của các bài kiểm tra từ vựng trong các nghiên cứu thực nghiệm trong các lĩnh vực kiểm tra ngôn ngữ và giảng dạy ngôn ngữ thứ hai, có thể kể đến các nghiên cứu về mức độ phù hợp của một bài kiểm tra với một mục đích giảng dạy cụ thể hay các nghiên cứu để hiểu rõ hơn một quá trình thu nhận từ vựng của người học ngôn ngữ thứ hai, không có bất kỳ quyết định nào về người học được đưa ra mà nghiên cứu chỉ để giải quyết các câu hỏi do các nhà nghiên cứu đưa ra; (3.) Công dụng trong đánh giá liên quan đến việc ra quyết định về chất lượng của các phương pháp và chương trình giảng dạy ngôn ngữ. Với công dụng này, các bài kiểm tra từ vựng có thể được đưa vào một loạt các biện pháp để đánh giá toàn bộ chương trình giảng dạy ngôn ngữ.

Cần phải lưu ý rằng một bài kiểm tra cụ thể có thể có nhiều hơn một công dụng. Việc đánh giá công dụng của một bài kiểm tra cần xem xét thông qua tính liên quan và mức độ hữu dụng trong việc đưa ra những quyết định mà bài kiểm tra hướng tới.

Tác động dự kiến và ảnh hưởng thực tế

Tác động, yếu tố thứ ba của mục đích kiểm tra, đề cập đến các tác động dự kiến của kiểm tra đối với cá nhân người học và giáo viên, các lớp học và chương trình đào tạo ngôn ngữ và thậm chí cả xã hội nói chung (Bachman & Palmer, 1996). Tác động có chủ đích cần được xem xét trong quá trình thiết kế bài kiểm tra trong khi các ảnh hưởng thực tế được đánh giá như là một phần của quá trình xác trị đề thi. Lý thuyết hiện hành và thực tiễn trong việc đánh giá bài kiểm tra đặt tầm quan trọng đáng kể đến tác động của kiểm tra (Frederiksen, 1984; Canale, 1987; Messick, 1989; Bachman & Palmer, 1996), và trong lĩnh vực kiểm tra ngôn ngữ nói riêng, các nhà nghiên cứu rất quan tâm đến các tác động của các bài kiểm tra lớn đến việc dạy và học (Alderson & Wall, 1993; Messick, 1996; Wall, 1997, được trích dẫn trong Read & Chapelle, 2001).



Hình 1.6: Khung kiểm tra từ vựng (Read & Chapelle, 2001)

Các yếu tố trung gian

Các yếu tố trung gian được Read và Chapelle (2001) liệt kê bao gồm xác định cấu trúc, tổng kết kết quả và báo cáo, trình bày đề kiểm tra.

Xác định cấu trúc

Có ba thuyết ảnh hưởng đến việc xác định cấu trúc kiểm tra được liệt kê là thuyết năng lực, thuyết hành vi và thuyết tương tác. Với thuyết năng lực, kết quả kiểm tra được quy về đặc điểm của thí sinh và do đó các yếu tố về bối cảnh không có vai trò quan trọng trong việc thiết kế đề kiểm tra để đánh giá một năng lực cụ thể. Các tác giả có đưa ra ví dụ trọng điểm đề kiểm tra là xác định lượng từ vựng mà thí sinh biết khi không có bất cứ gợi ý nào về bối cảnh để suy diễn ý nghĩa từ. Thuyết thứ hai, thuyết hành vi, lại nhấn mạnh vai trò trung tâm của các đặc điểm bối cảnh, do đó, cần xác định cụ thể bối cảnh. Các đề kiểm tra được thiết kế dựa vào các nhiệm vụ tích hợp giống với việc sử dụng ngôn ngữ đời thực, từ đó tạo ra điều kiện cho việc sử dụng từ vựng phù hợp. Thuyết thứ ba, thuyết tương tác, kiểm tra từ vựng như một năng lực trong một bối cảnh sử dụng cụ thể. Ví dụ xác định cấu trúc kiểm tra theo thuyết tương tác là kiểm tra lượng từ vựng ở dạng viết trong lĩnh vực toán học.

Tổng hợp kết quả và báo cáo

Tổng hợp kết quả và báo cáo là việc sử dụng điểm kiểm tra từ đó có thể đánh giá mức độ liên quan và tiện ích của đề kiểm tra, cũng như cách thức kết quả của người dự thi được báo cáo và lý giải. Kết quả kiểm tra có thể được tóm lược dưới dạng một điểm hoặc dưới dạng một hồ sơ chứa nhiều thành phần tùy thuộc vào các mục đích kiểm tra khác nhau. Một điểm số tóm lược khả năng ngôn ngữ của thí sinh sẽ là đủ khi kết quả kiểm tra được sử dụng để đưa ra quyết định, như đầu vào đại học hay tuyển dụng. Ngược lại, khi mục đích kiểm

tra là sự tiến bộ hay chẩn đoán để phục vụ giảng dạy trong lớp học hoặc chương trình ngôn ngữ, thông tin cụ thể hơn là cần thiết.

Trình bày đề kiểm tra

Trong quá trình xây dựng đề thi, cần phải xem xét đối tượng và cách thức trình bày đề kiểm tra. Người xây dựng đề cần lựa chọn cách thức của bài kiểm tra sao cho có thể thu hút đối tượng cụ thể, chẳng hạn như sinh viên, giáo viên, nhà nghiên cứu hay ban quản lý chương trình.

Thiết kế đề kiểm tra từ vựng và xác trị đề thi

Việc xác định cấu trúc, tóm lược và báo cáo kết quả, và trình bày đề kiểm tra đều tác động tới việc thiết kế đề kiểm tra, trong đó thuyết xác định cấu trúc, thuyết năng lực, thuyết hành vi hay thuyết tương tác có ảnh hưởng quan trọng đến hình thức của bài kiểm tra. Thuyết năng lực thường dẫn tới một bài kiểm tra từ vựng riêng lẻ (discrete) với các câu hỏi độc lập với ngữ cảnh (context-independent), trong khi thuyết hành vi dẫn tới thiết kế bài kiểm tra có các nhiệm vụ (task), trong đó có đánh giá bao chứa từ vựng (embedded). Thuyết tương tác sẽ đánh giá từ vựng trong mối quan hệ với bối cảnh sử dụng cụ thể (context-dependent).

Các yếu tố này, được rút ra từ lý thuyết xác trị rồi sau đó lại định hướng cho quá trình xác trị, thông qua các câu hỏi hướng dẫn xác trị. Các vấn đề về độ giá trị cấu trúc tuân theo một cách logic từ ba thuyết đã nêu ở trên. Read và Chapelle (2001) đề xuất việc sử dụng các tiêu chí đánh giá tính hữu dụng của đề kiểm tra của Bachman và Palmer (1996) như một hướng dẫn để thực hiện việc phân tích độ giá trị. Việc đánh giá ba thành phần này - xác định cấu trúc, báo cáo kết quả và trình bày đề kiểm tra - rất hữu ích để xác định các gợi ý cụ thể, từ đó thực hiện xác trị.

Có thể đi đến kết luận rằng, khung kiểm tra từ vựng tiếng Anh của Read và Chapelle (2001) có nhiều nội dung kế thừa từ khung thiết kế kiểm tra ngôn ngữ của Bachnam và Palmer (1996), những nội dung riêng trong lĩnh vực kiểm tra từ vựng dựa trên những nghiên cứu về từ vựng tiếng Anh và kiểm tra từ vựng tiếng Anh là những đóng góp đáng ghi nhận. Vì vậy, khung thiết kế kiểm tra từ vựng của Read và Chapelle (2001) được áp dụng rộng rãi trong các nghiên cứu cũng như các hoạt động kiểm tra đánh giá kiến thức từ vựng tiếng Anh.

1.1.3.5. Đánh giá kiến thức từ vựng tiếng Anh của người học ngoại ngữ ở Việt Nam

Trong bối cảnh dạy và học ngoại ngữ Việt Nam, từ vựng luôn có được xem trọng trong chương trình giảng dạy tiếng Anh ở Việt Nam. Trong Chương trình giáo dục phổ thông môn Tiếng Anh được ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo, hệ thống kiến thức từ vựng được tổng hợp trong Bảng 1.4.

Bảng 1.4: Từ vựng trong Nội dung dạy học các cấp của Chương trình giáo dục phổ thông môn Tiếng Anh 2018

Cấp	Nội dung dạy học	Số lượng từ vựng	Bậc tương ứng trong Khung năng lực ngoại ngữ Việt Nam
Tiểu học	Những từ thông dụng, đơn giản	khoảng 600 - 700 từ.	Bậc 1
Trung học cơ sở	Những từ thông dụng được thể hiện trong hai	khoảng 800 - 1000 từ ở bậc 2 (không bao gồm	Bậc 2

Cấp	Nội dung dạy học	Số lượng từ vựng	Bậc tương ứng trong Khung năng lực ngoại ngữ Việt Nam
	lĩnh vực ngôn ngữ nói và ngôn ngữ viết	các từ đã học ở tiểu học).	
Trung học phổ thông	Những từ thông dụng được thể hiện trong hai lĩnh vực ngôn ngữ nói và ngôn ngữ viết liên quan đến các chủ điểm và chủ đề trong Chương trình.	khoảng 600 - 800 từ ở Bậc 3 (không bao gồm các từ đã học ở các cấp tiểu học và trung học cơ sở)	Bậc 3

Nếu so sánh với đặc tả về từ vựng được mô tả trong Khung năng lực ngoại ngữ Việt Nam theo Thông tư 01/2014/TT-BGDĐT với 2 khung về phạm vi từ vựng (vocabulary range) và mức kiểm soát từ vựng (vocabulary control) được trích lại trong Bảng 1.4, có thể thấy được nỗ lực Chương trình giáo dục phổ thông môn Tiếng Anh trong việc lượng hóa mục tiêu từ vựng của học sinh sau khi hoàn thành chương trình phổ thông, với con số là khoảng 2500 từ ở bậc 3, khoảng 1700 từ ở bậc 2 và khoảng 700 từ ở bậc 1. Ngoài việc định lượng số từ đích mà học sinh có thể đạt được sau các bậc học, các vấn đề về chiều sâu từ vựng còn chưa được cụ thể và chi tiết để có thể hướng dẫn cho công tác kiểm tra đánh giá. Danh sách từ nào, khía cạnh ngôn ngữ nào, hình thức và công cụ

kiểm tra đánh giá để xác định được mục tiêu vẫn chưa được xác định một cách cụ thể.

Bảng 1.5: Đặc tả về từ vựng theo các bậc năng lực ngôn ngữ trong Khung năng lực ngoại ngữ Việt Nam

Bậc	Phạm vi từ vựng	Kiểm soát từ vựng
Bậc 1	- Có vốn từ cơ bản gồm các từ, cụm từ đơn lẻ thuộc các tình huống cụ thể.	- Không có đặc tả tương ứng.
Bậc 2	- Có đủ vốn từ để thực hiện các giao dịch thường nhật liên quan đến các tình huống và chủ đề quen thuộc. Có đủ vốn từ để diễn đạt những nhu cầu giao tiếp cơ bản và để xử lý những nhu cầu tối giản.	- Có khả năng kiểm soát được vốn từ hẹp thuộc các nhu cầu cụ thể hằng ngày.
Bậc 3	- Có đủ vốn từ để diễn đạt bản thân, tuy còn vòng vo, về hầu hết các chủ đề liên quan đời sống hằng ngày như gia đình, thói quen, sở thích, công việc, du lịch và các sự kiện đang diễn ra.	- Kiểm soát tốt lượng từ vựng ở trình độ sơ cấp. Tuy vậy vẫn có những lỗi lớn khi diễn đạt những ý nghĩ phức tạp hay những chủ đề và tình huống không quen thuộc.
Bậc 4	- Có vốn từ rộng về các vấn đề liên quan tới lĩnh vực của người sử dụng ngôn ngữ và hầu hết các chủ đề chung. Có khả năng hành văn đa dạng nhằm tránh lặp từ thường xuyên, nhưng do vốn từ vẫn còn thiếu nên khi diễn đạt vẫn còn ngập ngừng, dài dòng.	- Mức độ chính xác trong việc sử dụng từ nhìn chung là cao. Tuy đôi chỗ còn gây hiểu nhầm và sự lựa chọn từ còn chưa chính xác, nhưng điều đó không làm cản trở quá trình giao tiếp.
Bậc 5	- Thông thạo một lượng từ vựng lớn, có thể cho phép lấp đầy những	- Đôi khi có những khiếm khuyết nhỏ nhưng không có những lỗi

Bậc	Phạm vi từ vựng	Kiểm soát từ vựng
	khoảng trống bằng cách sử dụng những lời nói giải thích dài dòng, phải tìm kiếm từ thích hợp hoặc dùng các lời nói lảng tránh. Thông thạo các cụm từ mang tính thành ngữ và từ ngữ thông tục.	nhghiêm trọng trong việc sử dụng từ.
Bậc 6	- Thông thạo một lượng từ vựng rất lớn bao gồm các cụm từ mang tính thành ngữ, từ ngữ thông tục, nhận biết được mức độ ý nghĩa biểu cảm.	- Sử dụng từ luôn chính xác và thích hợp.

1.2. Tổng quan nghiên cứu

1.2.1. Các nghiên cứu về kiểm tra thích ứng trong đào tạo ngôn ngữ

Việc áp dụng kiểm tra thích ứng trên máy tính (CAT – Computerized Adaptive Testing) bắt đầu từ những năm 1980. Larson và Madsen (1985) đã phát triển các bài kiểm tra đầu tiên có sự hỗ trợ của máy tính, trong đó học sinh được phép truy cập các nguồn tài liệu tham khảo với việc sử dụng máy tính để hoàn thành bài kiểm tra viết của mình. Sau đó, các bài kiểm tra có sự chuyển đổi từ trên giấy và bút chì sang các bài kiểm tra trên máy tính nhờ các chuyên gia máy tính (Meunier, 1994). Năm 1986, một dự án tại Đại học Brigham Young đã giới thiệu bài kiểm tra ngôn ngữ thích ứng trên máy tính đầu tiên, một hệ thống kiểm tra ngôn ngữ sử dụng máy tính để điều chỉnh các bài kiểm tra cho phù hợp với mức độ thông thạo ngôn ngữ của người dự thi (Larson, 1989).

Kể từ đó, nhiều bài kiểm tra ngôn ngữ thích ứng trên máy tính đã được phát triển trong nhiều nghiên cứu thực nghiệm cho các ngôn ngữ khác nhau,

bao gồm tiếng Pháp (Burston, 1995; Laurier, 1999); Tiếng Tây Ban Nha (Larson, 1987); Tiếng Nhật (Brown & Iwashita, 1996); Tiếng Đức (Starr-Egger, 2001); Tiếng Trung (Wang và cộng sự, 2012). Đáng chú ý là DIALANG, một dự án châu Âu, cung cấp các bài kiểm tra khả năng thích ứng dựa trên web cho các ngôn ngữ chính thức của EU cùng với tiếng Ireland, và tiếng Na Uy (Chalhoub – Deville & Deville, 1999).

CAT đã được áp dụng rộng rãi trong kiểm tra đánh giá tiếng Anh hướng tới các khía cạnh khác nhau của trình độ thông thạo tiếng Anh (Tseng, 2016), thường xuyên hơn với từ vựng (Aviad-Levitzky và cộng sự, 2019; Laufer & Goldstein, 2004; Mizumoto, Sasao & Webb, 2019; Tseng, 2016; Vispoel, 1993, 1998; Vispoel và cộng sự, 1994), và kỹ năng nghe đọc (Chalhoub-Deville, 1999; Dunkel, 1999; Gawliczek và cộng sự, 2021; He & Min, 2017; Madsen, 1991; Nogami & Hayashi, 2010), và ít thường xuyên hơn với các kỹ năng nói viết (Malabonga, 2000; Malabonga & Kenyon, 1999; Stevenson & Gross, 1991).

Giờ đây, ngày càng có nhiều bài kiểm tra ngôn ngữ thích ứng trên máy tính (CALT - Computerized Adaptive Language Testing) đã được phát triển; các tổ chức kiểm tra đánh giá lớn, mặc dù trước đây ít chú ý đến CALT, nhưng hiện đã mở đường cho các bài kiểm tra ngôn ngữ thích ứng (Pathan, 2012). Một trong số đó, Educational Testing Service (ETS), đã chuyển đổi các bài kiểm tra chuẩn hóa mang tính quyết định (high-stakes exams) như TOEFL, GRE, và GMAT sang CALT (Rudner, 2010), gần đây có thể kể đến Duolingo kiểm tra tích hợp nhiều kỹ năng trong một bài thi (Brenzel & Settles, 2017). Ngoài ra, việc áp dụng CAT trong đánh giá ngôn ngữ là trọng tâm thảo luận trong nhiều các ấn phẩm trong những thập kỷ qua (Alderson, 2007; Brown, 1997; Canale, 1986; Chalhoub-Deville & Deville, 1999; Chapelle & Douglas, 2006; Henning, 1987; Khoshsima & Toroujeni, 2017; Lange, 1990; Larson & Madsen, 1985;

Laurier, 2000; Meunier, 1994; Ockey, 2009; Okhotnikova và cộng sự, 2019; Pathan, 2012; Tung, 1986;).

Bằng cách xem xét cả các cuộc thảo luận trước đây và các nghiên cứu thực nghiệm về đề kiểm tra thích ứng bằng máy tính, phần này thảo luận các cơ hội và thách thức của CAT trong kiểm tra đánh giá ngôn ngữ, để từ đó đưa ra nhận định về tính khả thi của việc xây dựng bài trắc nghiệm thích ứng của luận án.

1.2.1.1. Lợi thế của kiểm tra ngôn ngữ theo hình thức trắc nghiệm thích ứng bằng máy tính

Bài kiểm tra ngôn ngữ thích ứng trên máy tính có những lợi thế vượt trội so với bài kiểm tra với định dạng cố định (fixed tests) thông thường, bao gồm kiểm tra ngôn ngữ trên giấy và kiểm tra trên máy tính. Phần này sẽ thảo luận các lợi thế của CALT về độ chính xác cao, khả năng tiết kiệm thời gian, trải nghiệm thi cử được cá nhân hóa và độ phủ rộng các khả năng kiểm tra đánh giá mà CALT cung cấp.

CAT có độ chính xác cao

Một trong những thành tố của CALT là ngân hàng câu hỏi được hiệu chuẩn trên lý thuyết hồi đáp IRT. Vì vậy, CALT hứa hẹn mức độ chuẩn hóa cao hơn so với các bài thi thông thường như kiểm tra trên giấy và kiểm tra dựa trên máy tính. Nhiều nghiên cứu thực nghiệm đã cung cấp bằng chứng về độ chính xác của CALT so với các công cụ kiểm tra khác. Mizumoto và cộng sự (2019) đã tiến hành một nghiên cứu để phát triển và đánh giá một bài kiểm tra thích ứng trên máy tính CAT-WPLT. Đây là phiên bản CAT của Bài kiểm tra mức độ hiểu biết thành tố từ (WPLT: Word Part Levels Test) được thiết kế hai năm trước đó bởi Sasao và Webb (2017). Kết quả nghiên cứu kết luận rằng

phiên bản CALT tạo ra “độ chính xác tương tự hoặc cao hơn so với phiên bản câu hỏi cố định” (tr. 120). Gawliczek và cộng sự (2021) gần đây đã thực hiện một nghiên cứu để phát triển một bài kiểm tra thích ứng cho kỹ năng đọc và nghe, trong đó nhóm đã thực hiện một phân tích so sánh giữa CALT và bài kiểm tra dựa trên giấy. Một trong những chỉ số về độ giá trị và độ tin cậy được trình bày trong kết quả nghiên cứu là tỷ lệ cao những người dự thi xác nhận khả năng đọc và nghe của họ được đánh giá chính xác dựa trên kết quả CALT.

CALT cũng được coi là một công cụ đo lường trong giáo dục với độ chính xác cao bởi nhiều học giả, trong đó có Choi và cộng sự, 2003; Giourogrou & Economides, 2003; He & Min, 2017; Khoshsima & Toroujeni, 2017; Larson & Madsen, 1985; Meunier, 1994; Olsen và cộng sự, 1989; Pathan, 2012.

CAT mang lại hiệu quả về thời gian

Vì CALT đánh giá trình độ thông thạo ngôn ngữ dựa trên các câu hỏi phù hợp với từng người dự thi, các câu hỏi không quá khó và quá dễ sẽ không được lựa chọn. Kết quả là, cùng với thời gian tiến hành kiểm tra, số lượng câu hỏi cần thiết cho mỗi bài thi cũng giảm đáng kể trong khi vẫn đạt được kết quả kiểm tra chính xác. Olsen và cộng sự (1989), trong nghiên cứu so sánh của họ về các bài kiểm tra trên giấy, bài kiểm tra dựa trên máy tính và bài kiểm tra thích ứng bằng máy tính, đã phân tích thời gian làm bài và báo cáo rằng CAT chỉ mất một phần tư thời gian làm bài so với các bài kiểm tra trên giấy và khoảng một phần ba đến một nửa so với các bài kiểm tra trên máy tính. Madsen (1991) cũng kết luận rằng đa số người dự thi CALT (trên 80%) cần ít hơn một nửa số câu hỏi thi so với các bài kiểm tra trên giấy. Mizumoto và cộng sự (2019) báo cáo rằng phiên bản CAT xác định trình độ ngôn ngữ của người dự thi trong khoảng 10 phút thay vì 20 đến 30 phút theo yêu cầu của phiên bản bài kiểm tra trên giấy.

Các nhà nghiên cứu khác cũng xác nhận số lượng câu hỏi thi và thời gian làm bài giảm đáng kể trong CALT (Giourogrou & Economides, 2003; Khoshsima & Toroujeni, 2017; Meunier, 1994; Okhotnikova và cộng sự, 2019; Pathan, 2012; Tseng, 2016; Vispoel, 1993). Do đó, lợi thế rõ rệt này của CALT và kiểm tra ngôn ngữ dựa trên máy tính giúp giảm bớt gánh nặng về thời gian và chi phí liên quan đến quá trình tổ chức thi (Pathan, 2012) và mang lại nhiều cơ hội dễ tiếp cận hơn cho những người dự thi ở những nơi khác nhau, ngay cả đối với những người khuyết tật (Stone & Davey, 2011).

CAT mang đến trải nghiệm thi cử tốt hơn

Trong bài kiểm tra ngôn ngữ thích ứng, câu hỏi đầu tiên được chọn từ một ngân hàng câu hỏi lớn trong khi các câu hỏi tiếp theo được lựa chọn dựa trên câu trả lời của thí sinh đối với câu hỏi trước đó; do đó, mỗi người dự thi sẽ nhận được một bài kiểm tra phù hợp với trình độ thông thạo ngôn ngữ của họ. Khi thí sinh được kiểm tra với các câu hỏi phù hợp về độ khó và độ phân biệt, họ có thể tập trung tốt hơn vào bài thi, có quan điểm và cảm nhận về thi cử tích cực hơn (Meunier, 1994; Wise, 2014), đồng thời giảm bớt căng thẳng, buồn chán và mệt mỏi (Giourogrou & Economides, 2003; Rasskazova và cộng sự, 2017). Gawliczek và cộng sự (2021) đã chỉ ra mức độ động lực cao hơn với những người dự thi CALT có trình độ ngoại ngữ cả cao và thấp. Hơn nữa, vì không có giới hạn về thời gian làm bài cho từng câu hỏi và toàn bộ bài thi, thí sinh có thể kiểm soát tốc độ trả lời của mình để hoàn thành bài thi. Nói cách khác, người dự thi không còn lo lắng và áp lực về thời gian, những yếu tố có thể ảnh hưởng tiêu cực đến kết quả bài thi (Meunier, 1994).

Sau khi bài kiểm tra kết thúc, người dự thi CALT có thể truy cập kết quả và phản hồi ngay lập tức, với mức độ thành thạo ngôn ngữ được xác định và các đề xuất để cải thiện trình độ (Burston & Neophytou, 2014; Dandonoli,

1989; Gawliczek và cộng sự, 2021; Meunier, 1994). Kết quả được đánh giá riêng theo mức độ thành thạo nhất định và được thông báo cho người dự thi, không liên quan đến thành tích của những người dự thi khác. Đối với những thí sinh nóng lòng muốn biết kết quả, CALT mang đến cho họ trải nghiệm rất tích cực so với các hình thức thi thông thường khác. Cũng cần lưu ý rằng những thí sinh khác sẽ không có khả năng nhận được những thông tin hữu ích liên quan đến nội dung thi để có cơ hội gian lận, kể cả giữa những thí sinh ngồi cạnh nhau hay những thí sinh làm bài thi sau đó. Do đó, cách tiếp cận này của CALT có thể đảm bảo tính bảo mật cao hơn (Meunier, 1994; Okhotnikova và cộng sự, 2019; Pathan, 2012; Rasskazova và cộng sự, 2017). Chính lợi thế này có thể lý giải cho tỉ lệ cao các thí sinh thể hiện họ sẵn lòng tham gia CALT như trong nghiên cứu của Gawliczek và cộng sự (2021).

CAT phục vụ đa dạng các mục đích kiểm tra đánh giá

Okhotnikova và cộng sự (2019) đã đề cập đến sự đa dạng các trình độ năng lực ngôn ngữ có thể được đánh giá trong CALT như là một trong những lợi thế lớn nhất của CALT bên cạnh các tính năng tích cực của một bài kiểm tra có sự hỗ trợ của máy tính. Rasskazova và cộng sự (2017) đã nhấn mạnh thế mạnh này của các bài kiểm tra ngôn ngữ thích ứng bằng cách so sánh với các bài kiểm tra truyền thống cố định trong đó mức độ thông thạo ngôn ngữ cao nhất và thấp nhất không được đánh giá chính xác nếu các bài kiểm tra nhằm mục tiêu đến những thí sinh có trình độ trung bình. Olsen và cộng sự (1989) khẳng định rằng CALT có thể được phát triển cho cả đánh giá tham chiếu tiêu chí (criterion-referenced assessment) và đánh giá tham chiếu chuẩn (norm-referenced assessment) để đáp ứng các mục tiêu giáo dục cụ thể. Các bài kiểm tra thích ứng cũng có thể được khai thác cho các mục đích đánh giá khác nhau, chẳng hạn như các bài thi chuẩn hóa (high-stakes exams) (Alderson, 2007; Chapelle & Voss, 2008; He & Min, 2017), đánh giá quy mô lớn (Chalhoub-

Deville, 1999; Khoshsima & Toroujeni, 2017; Pathan, 2012; Wen & Qinghua, 2002), đánh giá chẩn đoán (Alderson, 2007; Chapelle & Voss, 2008; Larson & Madsen, 1985), và đánh giá xếp lớp đầu vào (Choi & McClenen, 2020; Giouroglou & Economides, 2005).

Ngoài ra, việc sử dụng thiết bị di động trong công nghệ máy tính có thể tạo điều kiện tích hợp các tính năng đa phương tiện trong các bài kiểm tra, quản lý linh hoạt các bài kiểm tra và cung cấp các bài kiểm tra theo yêu cầu (Triantafillou và cộng sự, 2008). Tính ưu việt của CAT so với các phương thức kiểm tra khác dẫn đến việc ứng dụng không giới hạn trong các lĩnh vực khác nhau từ giáo dục và đào tạo đến phát triển chuyên môn cho giáo viên trong trường học, công sở và thậm chí là quân đội. Trong nền giáo dục 4.0, khi việc sử dụng máy tính phổ biến ở tất cả các gia đình và trường học, chắc chắn CALT có thể đáp ứng nhu cầu đánh giá ngày càng cao hơn với nhiều tùy chọn đa dạng và những nỗ lực nâng cấp liên tục.

1.2.1.2. Thách thức của kiểm tra ngôn ngữ theo hình thức trắc nghiệm thích ứng bằng máy tính

Dù CALT mang lại nhiều điểm tích cực cho người dự thi, các học giả đã đưa ra một số thách thức liên quan đến các mối lo ngại về kiểm tra đánh giá và các mối lo ngại về thiết kế. Trong kỷ nguyên chuyển đổi kỹ thuật số 4.0, một số thách thức không còn là tâm điểm chú ý trong khi những thách thức khác lại đòi hỏi những cân nhắc và hành động nghiêm túc để có trải nghiệm thi cử tích cực hơn với CALT.

Những quan ngại về kiểm tra đánh giá

Trong kỷ nguyên kỹ thuật số, khi máy tính trở thành thiết bị quen thuộc trong giáo dục, kiến thức chuyên môn máy tính (Madsen, 1991), việc trình

chiều câu hỏi với giao diện máy tính và những thiên lệch trong đánh giá do thí sinh thiếu kỹ năng máy tính (Henning, 1987) đã không còn là nguyên nhân chính gây ra thách thức về CALT. Thay vào đó, các mối lo ngại khác về tính khách quan công bằng trong kiểm tra nên được đưa ra thảo luận để rút ra một số đề xuất mang tính xây dựng.

Henning (1987) lo ngại rằng độ giá trị của CALT có thể bị ảnh hưởng bởi thời gian và tốc độ làm bài khác nhau, từ đó liên quan đến tính công bằng của thi cử. Một mối bận tâm khác về trải nghiệm với CALT là thí sinh không có cơ hội quay lại các câu hỏi trước đó để hoàn thành và sửa lại câu trả lời (Giourogrou & Economides, 2003; Khoshsimā & Toroujeni, 2017). Khi thí sinh trả lời câu hỏi, năng lực của thí sinh được tính để hệ thống lựa chọn câu hỏi tiếp theo. Thuộc tính này của CALT có thể thách thức thói quen kiểm tra của một số thí sinh. Tuy nhiên, những lo ngại này có thể được khắc phục khi thí sinh có thể tiếp cận dễ dàng hơn và quen thuộc hơn với CALT.

Wainer và Eignor (2000) cũng nêu lên quan ngại của họ về tính bảo mật của CALT khi một số hạng mục quan trọng có thể được ghi nhớ và thảo luận giữa các thí sinh, điều này ảnh hưởng đến độ tin cậy của CALT. Để vượt qua thách thức này, Mizumoto và cộng sự (2019) đưa ra đề xuất về một quy trình phát triển đề thi nghiêm ngặt và việc xây dựng ngân hàng câu hỏi với số lượng lớn câu hỏi thi.

Những quan ngại về quá trình thiết kế

Các bài báo về khung thiết kế, hướng dẫn và kinh nghiệm chia sẻ đều chỉ ra rằng quá trình phát triển của một hệ thống kiểm tra thích ứng đòi hỏi một sự đầu tư rất lớn các nguồn lực như thời gian, tiền bạc, sức lực cũng như trí tuệ con người, cụ thể là chuyên môn liên ngành liên quan đến kiểm tra đánh giá ngôn ngữ và công nghệ máy tính (Brown, 1997; Chapelle, 1999; Chapelle &

Voss, 2008; Chen & Wang, 2010; Larson, 1998; Nydick & Weiss, 2009; Pathan, 2012; Rasskazova và cộng sự, 2017; Spoden, Frey & Bernhardt, 2018).

Liên quan đến tính đơn chiều của các mô hình IRT được sử dụng trong thiết kế các bài kiểm tra thích ứng, nhiều học giả đã thể hiện nhiều mối bận tâm (Chalhoub-Deville, 2010; Jamieson, 2005; Liu, 2019). Canale (1986) lo ngại về hạn chế của CALT đối với đánh giá khả năng ngôn ngữ khi khả năng ngôn ngữ bao gồm nhiều khía cạnh như nhận thức, kiến thức, kỹ năng sử dụng ngôn ngữ theo ngữ cảnh. Giouroglou và Economides (2003), He và Min (2017), Norris (2001), Okhotnikova và cộng sự (2019), và Young và cộng sự (1996) bày tỏ lo ngại của họ về các giới hạn về tính đơn chiều của CALT trong việc xử lý câu hỏi mở và đặt nghi vấn về tính khả thi của CALT trong việc đánh giá các kỹ năng như viết nói. Ngay cả khi một số hệ thống chấm điểm tự động đã được phát triển, hiện vẫn tồn tại nhiều tranh cãi về việc liệu có phải chỉ người thực mới có thể đánh giá được kỹ năng viết và nói hay không. Có thể thấy rằng việc phát triển CALT đòi hỏi nhiều nỗ lực hơn nữa, đặc biệt là việc phát triển hệ thống chấm điểm tự động, để xử lý những thách thức của việc đánh giá kỹ năng nói và viết để làm cho CALT có tính tương tác hơn và đáng tin cậy như mong đợi trong triết lý giáo dục nói chung, trong đào tạo ngôn ngữ hiện đại, cũng như thách thức về áp dụng công nghệ và thực tiễn kiểm tra đánh giá (Canale, 1986; Okhotnikova và cộng sự, 2019; Pathan, 2012; Young và cộng sự, 1996).

Việc thiết kế hệ thống CALT bao gồm hai phần không thể tách rời: ngân hàng câu hỏi đã chuẩn hóa và các thuật toán CAT. Người ta tin rằng một ngân hàng lớn các câu hỏi được chuẩn hóa có thể đảm bảo chất lượng nội dung và sự vận hành thành công của hệ thống CALT. Okhotnikova và cộng sự (2019) đề xuất rằng ngân hàng câu hỏi của CALT nên bao gồm hơn một nghìn câu hỏi để đảm bảo đánh giá được tất cả các mức độ kết quả. Meunier (1994) chỉ ra hầu

hết các hệ thống kiểm tra thích ứng có số lượng 100 hoặc 200 câu hỏi được chuẩn hóa với mô hình IRT một tham số và khuyến nghị sử dụng 2000 câu hỏi trở lên trong trường hợp mô hình IRT hai và ba tham số được sử dụng để chuẩn hóa câu hỏi. Nydick và Weiss (2009) cũng như Suvorov và Hegelheimer (2013) đều nhấn mạnh sự cần thiết của ngân hàng câu hỏi lớn để giảm thiểu các vấn đề về gian lận phát sinh từ việc các thí sinh có thể ghi nhớ các câu hỏi. Mặc dù nhiều nghiên cứu thực nghiệm đã được thực hiện nhưng vẫn chưa có kết luận về số lượng câu hỏi tối thiểu của ngân hàng câu hỏi trong một hệ thống CALT; do đó, yêu cầu đặt ra là các nghiên cứu về CALT để đưa ra các chỉ số tiêu chuẩn của ngân hàng câu hỏi của các dự án thiết kế CALT trong tương lai.

Về thách thức thiết kế liên quan đến thuật toán CAT, hiện vẫn chưa có các quy định cụ thể về điểm bắt đầu, lựa chọn câu hỏi, cơ chế tính điểm, và tiêu chuẩn kết thúc. Các nhà nghiên cứu vẫn đang tiến hành các nghiên cứu tiếp theo để xác định những yêu cầu tối thiểu mà có thể tối đa hóa tiềm năng của CALT trong việc cung cấp các phương pháp đánh giá ngôn ngữ đáng tin cậy và hiệu quả (Khoshsima & Toroujeni, 2017; Pathan, 2012). Bên cạnh đó, trong thời đại 4.0 của chuyển đổi kỹ thuật số, khi công nghệ máy tính có những bước phát triển vượt bậc nhờ vào các tiến bộ kỹ thuật, thu hút mọi quan tâm ngày càng tăng trong mọi lĩnh vực, được ứng dụng ngày một rộng rãi và toàn diện trong các hệ thống giáo dục cũng như trong các hoạt động đánh giá ngôn ngữ, công nghệ máy tính không chỉ mang lại nhiều cơ hội để phát triển CALT, mà còn mang đến những thách thức không hề nhỏ liên quan đến việc xử lý các nguy cơ về an ninh mạng và đáp ứng các yêu cầu nghiêm ngặt về duy trì, quản lý và nâng cấp hệ thống CALT cũng như sửa đổi các thuật toán để đáp ứng nhu cầu ngày càng cao của xã hội đối với giáo dục nói chung và kiểm tra đánh giá nói riêng. Những thách thức này có thể được khắc phục thông qua sự đầu tư

nguồn lực phù hợp và sự hợp tác từ các lĩnh vực khác nhau, bao gồm dạy và học ngôn ngữ, kiểm tra đánh giá, lập trình và quản lý hệ thống máy tính.

Xét trong bối cảnh thực hiện luận án, mặc dù CAT còn khá mới mẻ ở Việt Nam, nhưng những tiền đề phát triển và báo cáo kết quả tích cực của hệ thống trắc nghiệm đã được xây dựng và hoàn thiện chắc chắn là những lợi thế đáng kể, giảm đi các áp lực liên quan việc thiết kế và vận hành hệ thống. Việc xây dựng bài trắc nghiệm thích ứng đánh giá từ vựng của luận án là khả thi, được xem như một bước mở rộng nội dung kiểm tra để phát triển hệ thống, khẳng định khả năng ứng dụng của CAT trong đào tạo ngôn ngữ cũng như trong lĩnh vực kiểm tra đánh giá ở Việt Nam.

1.2.2. Các nghiên cứu về đánh giá kiến thức từ vựng tiếp nhận tiếng Anh

Theo Meara (1987), từ vựng trở lại như một lĩnh vực được quan tâm trong những năm 1980-1990, kể từ đó cả lĩnh vực giảng dạy và nghiên cứu chứng kiến sự gia tăng của các bài kiểm tra từ vựng. Có thể thấy rõ sự gia tăng này thông qua một phần danh sách các bài kiểm tra độ rộng từ vựng: bài kiểm tra từ vựng EFL (Meara, 1992) và tất cả các phiên bản, ví dụ AuralLex (Milton & Hopkins, 2006), bài kiểm tra cấp độ từ vựng Vocabulary Levels Test (VLT) (Nation, 1983; Schmitt và cộng sự, 2001), bài kiểm tra cấp độ từ vựng mới New Vocabulary Levels Test (McLean & Kramer, 2015), bài kiểm tra cấp độ từ vựng cập nhật Updated Vocabulary Levels Test (Webb, Sasao, & Ballance, 2017), bài kiểm tra cấp độ từ vựng sản sinh Productive Vocabulary Levels Test (Laufer & Nation, 1999), bài kiểm tra mức độ từ vựng Nghe Listening Vocabulary Levels Test (McLean và cộng sự, 2015), Lex30 (Meara & Fitzpatrick, 2000), bài kiểm tra từ vựng dành cho người học tiếng Anh nâng cao Lexical Test for Advanced Learners of English (Lemhöfer & Broersma, 2012), bài kiểm tra thích ứng bằng máy tính về độ rộng và mức độ hiểu biết từ vựng Computer

Adaptive Test of Size and Strength (CATSS) (Laufer & Goldstein, 2004; Aviad-Levitzky, Laufer & Goldstein, 2019), bài kiểm tra kích thước từ vựng Vocabulary Size Test (VST) rất phổ biến (Nation & Beglar, 2007), với rất nhiều phiên bản và chuyển thể song ngữ (Coxhead và cộng sự, 2015; Nguyen & Nation, 2011).

Dù các bài kiểm tra từ vựng được phát triển rất đa dạng, có các cách tiếp cận khác nhau, không thể phủ nhận rằng kiểm tra từ vựng tiếp nhận tiếng Anh là phổ biến nhất. Hiện nay, nhiều nghiên cứu về kiểm tra từ vựng của người học tiếng Anh tại Việt Nam đều sử dụng các bài kiểm tra từ vựng tiếp nhận dạng viết để xác định vốn từ vựng của thí sinh, có thể kể đến các nghiên cứu kiểm tra từ vựng của sinh viên đại học của Dang (2020), Le và Nation (2011), Nguyen và Webb (2017) và các nghiên cứu kiểm tra từ vựng của học sinh phổ thông của Nguyen (2021) và Vu và Nguyen (2019).

Với mục đích tìm kiếm bài kiểm tra kiến thức từ vựng tiếp nhận dạng viết hướng tới đối tượng người học tiếng Anh ở Việt Nam, bốn bài kiểm tra được xây dựng và phát triển gần đây sẽ được tổng lược vì có phạm vi sử dụng rộng rãi và phù hợp với xu hướng áp dụng công nghệ máy tính bao gồm: 1) bài kiểm tra cấp độ từ vựng Vocabulary Levels Test (VLT); 2) bài kiểm tra kích thước từ vựng Vocabulary Size Test (VST); 3) bài kiểm tra thích ứng bằng máy tính về độ rộng và mức độ hiểu biết từ vựng Computer Adaptive Test of Size and Strength (CATSS) và 4) bài kiểm tra từ vựng thông dụng New General Service List Test (NGSLT).

1.2.2.1. Bài kiểm tra cấp độ từ vựng VLT

Bài kiểm tra cấp độ từ vựng (VLT) là bài kiểm tra từ vựng được công bố từ 1983 và được sử dụng rộng rãi nhất trong việc giảng dạy và nghiên cứu. VLT là một công cụ để đo lường từ vựng dạng viết của thí sinh trong năm nhóm từ

vựng. Các nhóm 2.000 từ đầu tiên lấy danh sách từ General Service List GSL của West (1953), và các nhóm 3.000, 5.000 và 10.000 từ lấy các từ trong danh sách của Thorndike và Lorge (1944) và Kucera và Francis (1967) (như được trích dẫn trong McLean & Kramer, 2015). Hiện tại, các bài kiểm tra cấp độ từ vựng đều có thể truy cập được bao gồm VLT 2k-10k của Nation (1990), VLT 1k của Nation (2001), VLT v.2 của Schmitt và cộng sự (2001) và Updated VLT của Webb và cộng sự (2017).

Mục đích chính của bài kiểm tra là để chẩn đoán mức độ thông thạo từ vựng thông dụng của người học để chỉ định tài liệu học tập phù hợp. Mức độ thành thạo được xác định trên cơ sở trả lời đúng 29 câu hỏi hoặc nhiều hơn cho 30 câu hỏi tương ứng, vì cần có độ bao phủ 98% để đọc hiểu các tài liệu viết (Nation, 2013). Là một bài kiểm tra chẩn đoán, VLT được chứng minh là rất hữu ích trong việc hướng sự quan tâm đến mức độ từ vựng tương ứng của người dạy và người học. Các câu hỏi của VLT được thiết kế với cùng một định dạng ghép nối, trong đó người học phải ghép ba trong số sáu từ ở bên trái với các nghĩa được liệt kê ở bên phải.

1. apply		
2. elect	<input type="checkbox"/>	choose by voting
3. jump	<input type="checkbox"/>	become like water
4. manufacture	<input type="checkbox"/>	make
5. melt		
6. threaten		

Hình 1.7: Ví dụ câu hỏi trong VLT

Sau khi hoàn thành câu trả lời, trang web sẽ cho điểm về tỷ lệ phần trăm câu trả lời đúng trong số các từ được kiểm tra và đưa ra gợi ý về các nhóm vấn đề nếu câu trả lời sai ở các cấp độ.

Các phiên bản khác nhau đã được áp dụng rộng rãi trong cả đánh giá và nghiên cứu, và khá nhiều bài báo đã được xuất bản để xác định bài kiểm tra. Read (1988) thấy bài kiểm tra VLT phiên bản 1983 đáng tin cậy. Schmitt và cộng sự (2001) kết hợp bốn phiên bản gốc thành hai phiên bản, mỗi phiên bản có 10 cụm trong năm phần. Hai phiên bản này được kiểm tra bởi tổng cộng 801 đối tượng thuộc 13 nhóm ở năm quốc gia. Các chỉ số độ tin cậy (Cronbach's alpha) cho tất cả các phần tương ứng các cấp độ đều cao, nằm trong khoảng từ .915 (đối với phần 10k) đến .958 (đối với phần từ học thuật). Bài kiểm tra thông qua phân tích nhân tố được kết luận là đơn chiều; các cuộc phỏng vấn cá nhân cũng chỉ ra rằng những người được kiểm tra đón nhận bài kiểm tra và rằng các câu trả lời trong bài kiểm tra có thể phản ánh kiến thức từ vựng cơ bản của họ. Ngoài ra bài kiểm tra được cho là phân biệt tốt giữa người học giỏi hơn và người học yếu hơn và do đó bài kiểm tra có thể được sử dụng như bài kiểm tra xếp lớp cũng như để xếp học sinh nhanh chóng vào các nhóm năng lực dựa trên kiến thức từ vựng.

Tuy nhiên, định dạng câu hỏi của VLT trong các phiên bản trước đó được các học giả cho rằng có một số khuyết điểm: a) thiếu tính độc lập của các câu hỏi, b) sự không chính xác tương đối của định dạng khi so với một câu hỏi có bốn lựa chọn tiêu chuẩn, c) học sinh khó hiểu định dạng và d) khó điều chỉnh các bài kiểm tra với các hình thức kiểm tra khác hoặc kho ngữ liệu nguồn (Culligan, 2015; Kamimoto, 2014; McLean và Kramer, 2015). Đây cũng là lý do định dạng câu hỏi riêng lẻ cho mỗi từ với bốn lựa chọn được sử dụng trong phiên bản bài kiểm tra VLT mới năm 2015.

1. time: They have a lot of **time**.
a. money
b. food
c. hours
d. friends

Hình 1.8: Ví dụ câu hỏi trong New VLT

Ngoài vấn đề về định dạng câu hỏi, VLT cũng thiếu những bằng chứng xác trị như Schmitt và cộng sự (2020) đã chỉ ra: VLT phiên bản 1983 chỉ có một đoạn ngắn đề cập đến việc tiến hành kiểm tra thử và kết luận bài kiểm tra không nên được sử dụng với những người học có tiếng mẹ đẻ là ngôn ngữ Latin; VLT phiên bản 2001 có đề cập một nghiên cứu với 801 người có ngôn ngữ mẹ đẻ khác nhau và một số bằng chứng xác thực, nhưng không bàn về cách thức bài kiểm tra nên được sử dụng với những người học khác nhau hoặc trong các ngữ cảnh khác nhau; bài báo về VLT phiên bản 2015 cũng không tập trung vào vấn đề xác trị. Thêm vào đó, VLT cũng không có thông tin về ý nghĩa của điểm kiểm tra liên quan đến những gì người học thực sự có thể làm với các từ được kiểm tra (ngoài việc trả lời chúng trong bài kiểm tra) trong khi thông tin này rất có giá trị với người học và đáp ứng một mức độ xác thực nhất định của bài kiểm tra.

1.2.2.2. Bài kiểm tra độ rộng từ vựng VST

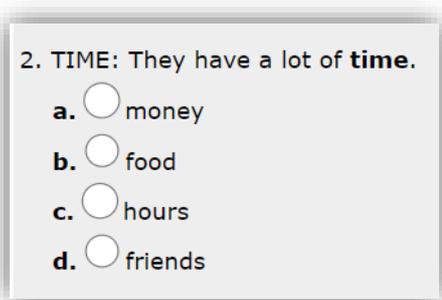
VST là một bài kiểm tra gần đây hơn về kích thước từ vựng của Nation và Beglar (2007). VST được thiết kế để đo lường kiến thức từ vựng tiếp nhận ở dạng viết, ước tính tổng số lượng từ vựng của thí sinh. Đây là một bài kiểm tra trình độ nhưng ước tính cũng có thể được sử dụng để chẩn đoán sự tăng trưởng vốn từ vựng dài hạn và đặt mục tiêu học từ vựng mới. VST có hai phiên bản, một phiên bản 14.000 chứa 140 câu hỏi trắc nghiệm và hai phiên bản

20.000 song song chứa 100 câu hỏi trắc nghiệm. Thí sinh có thể truy cập cả phiên bản kiểm tra giấy và phiên bản kiểm tra với máy tính cũng như một số phiên bản song ngữ của bài kiểm tra kích thước từ vựng VST đã được phát triển bao gồm tiếng Ả Rập, tiếng Gujarati, tiếng Nga, tiếng Hàn, tiếng Việt, tiếng Quan thoại và tiếng Nhật. Trong các phiên bản song ngữ, các từ đích là tiếng Anh và các lựa chọn sử dụng tiếng mẹ đẻ của người học. Các phiên bản song ngữ được nhận định gia tăng giá trị của bài kiểm tra từ vựng tiếp nhận khi vẫn kiểm tra đánh giá kiến thức từ vựng tiếp nhận mà giảm thiểu các yếu tố nhiễu như ngữ pháp hay kỹ năng đọc hiểu tiếng Anh của thí sinh (Le & Nation, 2011).

Bài kiểm tra trình bày từ được kiểm tra bằng một câu đã được loại bỏ ngữ cảnh cùng với bốn phương án thay thế ý nghĩa. VST sử dụng định dạng câu hỏi trắc nghiệm với bốn phương án lựa chọn nghĩa của từ được kiểm tra. Tổng điểm cần được nhân với 100 đối với phiên bản 14.000 và 200 đối với phiên bản 20.000 để tính tổng lượng từ vựng của thí sinh. Kết quả báo cáo tỷ lệ phần trăm kiến thức về từ của người học ở mỗi cấp độ (kiến thức một phần cũng được đánh giá ở đây) và tổng số lượng từ vựng ước tính và do đó bài kiểm tra có thể được sử dụng như một bài kiểm tra chẩn đoán độ rộng từ vựng tiếp nhận. Karami (2012) và Le và Nation (2011) đề xuất rằng mọi cấp độ của bài kiểm tra nên được bao phủ để tránh việc độ rộng từ vựng bị đánh giá thấp đi nhiều so với thực tế và có được ước tính hợp lệ hơn về độ rộng từ vựng của người học, mặc dù một số nghiên cứu khác ủng hộ rằng học sinh chỉ cần làm kiểm tra hai mức trên khả năng của họ (McLean & Kramer, 2015) để có độ chính xác kết quả ngày càng cao.

Về xác trị bài kiểm tra, bài báo ban đầu không có thông tin xác thực, nhưng một nghiên cứu tiếp theo của Beglar (2010) cung cấp bằng chứng hợp lệ cho VST bằng cách xác định độ giá trị cấu trúc về nội dung, cơ sở lý thuyết, cấu trúc, độ tin cậy từ khung xác trị của Messick và nghiên cứu khả năng đáp

ứng và khả năng diễn giải. Phương pháp tiếp cận dựa trên Rasch được áp dụng để đánh giá tính đơn chiều của bài kiểm tra. Có thể thấy rằng VST có thể phân biệt người học với các mức độ khác nhau của kiến thức từ vựng tiếp thu dạng viết và cung cấp đủ cấp độ từ để đo lường khả năng tiếp thu từ vựng của người học trong một khoảng thời gian dài. Riêng với phiên bản song ngữ tiếng Việt, thông tin xác trị được đưa ra với lượng thí sinh khá nhỏ, chỉ 62 sinh viên chuyên ngữ năm thứ 3 tại một trường đại học (Le & Nation, 2011). Schmitt và cộng sự (2020), dựa vào thông tin tổng hợp về các phiên bản song ngữ, đã đưa ra đề xuất về việc tiến hành thêm các nghiên cứu cung cấp thêm bằng chứng xác trị hợp lệ với các phiên bản song ngữ, trong đó có phiên bản song ngữ tiếng Việt.



Hình 1.9: Ví dụ câu hỏi VST

Ngoài ra, các nhà nghiên cứu cũng lưu ý việc suy luận thận trọng khi sử dụng các kết quả từ VST để đo kiến thức từ vựng của người học (Schmitt và cộng sự, 2020). Vì VST lấy mẫu mỗi 10 từ từ các nhóm 1.000 từ phổ biến nhất (lên đến nhóm thứ 14 hoặc 20 tùy thuộc vào phiên bản), “điểm của thí sinh cần được nhân với 100 để có tổng lượng từ vựng của họ” (Nation, 2013, tr. 525). Ví dụ, điểm 30 trên 140 sẽ tạo ra ước tính kích thước của 3.000 họ từ đã biết. Có thể thấy suy luận này khá đơn giản, dễ thực hiện và dễ hiểu; tuy nhiên, có hai vấn đề cần lưu ý ở đây, đó là tính đại diện của 10 từ lấy mẫu trong đề kiểm tra và việc sử dụng định dạng câu hỏi trắc nghiệm 4 lựa chọn mang lại 25% khả năng thí sinh trả lời đúng câu hỏi nhưng không thực sự biết từ kiểm tra. Để

giảm sai số trong kiểm tra từ vựng của thí sinh, các tác giả đưa ra đề xuất về việc sử dụng các mô hình lý thuyết hồi đáp IRT nhiều tham số hay áp dụng kiểm tra thích ứng với các ứng dụng công nghệ trong quá trình xây dựng các bài kiểm tra từ vựng.

1.2.2.3. Bài kiểm tra từ vựng thích ứng CATSS

Bài kiểm tra từ vựng thích ứng bằng máy tính CATSS hướng tới nhiều mục đích theo Schmitt và cộng sự (2020) tổng kết: 1) cung cấp cho các nhà nghiên cứu và giáo viên một hồ sơ chẩn đoán về mức độ phát triển ngôn ngữ từ vựng hiện tại của người học; 2) cung cấp một phương tiện đo lường sự phát triển cả về độ rộng và mức độ hiểu biết khi tiếp nhận từ vựng; 3) cho phép xếp lớp và nhập học hiệu quả hơn trong các chương trình giảng dạy ngôn ngữ vì kiến thức từ vựng đã được chứng minh là có liên quan đến sự thành công trong việc đọc, viết và thông thạo ngôn ngữ nói chung cũng như thành tích học tập.

Yếu tố độ rộng hay kích thước từ vựng của CATSS được phản ánh trong các mẫu từ từ năm nhóm: 1 - nhóm 2000 từ phổ biến nhất, 2 - nhóm 1000 từ phổ biến thứ ba, 3 - nhóm 1000 từ phổ biến thứ năm, 4 - nhóm 1000 từ phổ biến thứ mười và 5- nhóm từ vựng học thuật AWL (Academic Word List), cắt ngang các từ từ nhóm 1000 từ phổ biến thứ ba và thứ năm và cũng bao gồm một số từ ít thường xuyên hơn. Ba mươi câu hỏi được chọn ngẫu nhiên từ mỗi nhóm. Các câu hỏi này được lấy từ VLT của Schmitt và cộng sự (2001). Trong quá trình xác trị, đề xuất được đưa ra là 30 từ là đủ để được coi là đại diện cho mỗi nhóm bao gồm 1000 từ.

Yếu tố mức độ hiểu biết được phản ánh trong bốn mức độ hiểu biết cho mỗi từ, bắt đầu với mức khó nhất đến mức dễ nhất tương ứng với các yêu cầu ở mức khó nhất – cung cấp từ với nghĩa cho trước đến yêu cầu ở mức dễ nhất – chọn đáp án chứa nghĩa đúng nhất trong các đáp án cho sẵn. Do đó, việc cho

điểm của các mức độ hiểu biết khác nhau trong phiên bản CATSS phản ánh độ khó khác nhau này: 1 điểm cho câu trả lời đúng ở mức khó nhất; 0,75 điểm và 0,5 với hai mức tiếp theo và 0,25 điểm với mức dễ nhất.

Bảng 1.6: Định dạng câu hỏi trong CATSS

Nội dung kiến thức	Tính điểm	Yêu cầu của câu hỏi	Ví dụ
Productive recall	1	Cung cấp từ có nghĩa cho trước	The task is to supply the L2 target word (melt). <i>Turn into water</i> <i>m_____</i>
Receptive recall	0,75	Điền từ phù hợp vào một cụm từ hoặc một câu ngắn để giải thích nghĩa của từ cho trước	The task is to demonstrate understanding of the meaning of the L2 word (melt). <i>When something melts, it turns into _____.</i>
Productive recognition	0,5	Chọn đáp án đúng chứa từ từ có nghĩa cho trước trong số bốn phương án a-b-c-d.	The task is to choose the target word from four options (a b c d). <i>Turn into water</i> <i>a. elect</i> <i>b. blame</i> <i>c. melt</i> <i>d. threaten</i>
Receptive recognition	0,25	Chọn đáp án có giải thích đúng nhất nghĩa của từ cho trước trong bốn	The task is to choose the meaning of the target word from the four options provided. <i>When something melts, it _____</i>

Nội dung kiến thức	Tính điểm	Yêu cầu của câu hỏi	Ví dụ
		phương án 1-b-c-d.	<i>a. chooses</i> <i>b. accuses</i> <i>c. makes threats</i> <i>d. turns into water</i>

So với các đề kiểm tra từ vựng được tóm lược, CATSS là bài kiểm tra từ vựng duy nhất sử dụng hình thức thích ứng bằng máy tính. Với hình thức này, nếu thí sinh nhận được một câu hỏi về một từ với một mức độ hiểu biết nhất định, tùy câu trả lời của thí sinh, thí sinh sẽ nhận câu hỏi tiếp về từ khác hoặc về từ đó nhưng với các mức độ hiểu biết tương thích. Ví dụ: nếu thí sinh đã cung cấp từ có nghĩa cho trước (mức khó nhất), từ đó sẽ không xuất hiện lại trong ba mức còn lại. Khi không có phản hồi hoặc phản hồi không chính xác, máy tính sẽ giữ từ đó trong bộ nhớ của nó để kiểm tra từ đó ở mức dễ hơn, nhưng chỉ sau khi tất cả các từ khác ở cùng nhóm đã được kiểm tra ở mức đó. Một khi thí sinh trả lời đúng, từ đó sẽ không được kiểm tra lại với các mức thấp hơn.

Cần lưu ý rằng trong bài kiểm tra CATSS, từ vựng tiếp nhận chỉ là một nội dung kiểm tra đánh giá. CATSS cung cấp một bức tranh toàn diện hơn cả về độ rộng và mức độ hiểu biết từ vựng của thí sinh khi CATSS không chỉ tính đến số lượng từ mà người dự thi biết mà còn tính đến mức độ hiểu biết của những từ này (Laufer và cộng sự, 2004; Laufer & Goldstein, 2004). Tuy nhiên có một số thách thức với CATSS không thể phủ nhận được. Về định dạng của câu hỏi, có bốn dạng tương ứng với bốn mức độ hiểu biết từ vựng. Mặc dù hình thức thích ứng bằng máy tính được áp dụng, mỗi từ chỉ kiểm tra một mức độ hiểu biết, số lượng câu hỏi của bài kiểm tra là 140 câu có thể được xem là khá dài, cùng với định dạng khác nhau cho các câu hỏi có thể mang lại những khó

khẩn không nhỏ đối với thí sinh. Về định dạng câu hỏi với mức độ hiểu biết cao nhất, thí sinh được yêu cầu cung cấp dạng từ, chính những học giả thiết kế CATSS cũng ghi nhận đó là một thách thức lớn vì câu trả lời của thí sinh đưa ra có sự khác biệt lớn phụ thuộc vào trình độ của từng thí sinh, việc chuẩn bị đầy đủ đáp án để tiến hành chấm tự động trên máy tính đòi hỏi nỗ lực rất lớn của những nhà thiết kế cùng với những chuyên gia là người bản địa và những người có năng lực tương đương với thí sinh.

Ngoài ra, CATSS ban đầu sử dụng các từ được lấy mẫu không theo trình tự các nhóm từ - K1 + K2, K3, K5, K10 và AWL. Do đó, điều này có thể hạn chế việc đánh giá tổng lượng từ vựng của người học (McLean & Kramer, 2015). Đây cũng là lý do mà CATSS mới bản 2019 (Aviad-Levitzky và cộng sự, 2019) khắc phục hạn chế này bằng cách kiểm tra các từ từ 14 nhóm từ tuần tự theo cách tương tự như trong VST của Nation và Beglar (2007).

Có thể thấy rằng CATSS có những ưu thế khi áp dụng hình thức trắc nghiệm thích ứng bằng máy tính cùng với các bằng chứng xác thực đáng tin cậy, nhưng việc đáp ứng nhiều mục đích kiểm tra với việc đưa vào nhiều nội dung kiểm tra cùng nhiều định dạng câu hỏi có thể mang lại những ảnh hưởng không nhỏ đến trải nghiệm cũng như kết quả thể hiện của các thí sinh.

1.2.2.4. Bài kiểm tra từ vựng thông dụng NGS LT

NGSLT là một bài kiểm tra chẩn đoán về kiến thức tiếp nhận từ vựng tiếng Anh với danh sách từ thông dụng New General Service List (NGSL) (Stoeckel & Bennett, 2015). Bài kiểm tra có 100 câu hỏi, 20 câu hỏi cho mỗi dải trong số năm dải từ khoảng 560 từ của danh sách NGSL. Việc chia thành năm dải từ như vậy được giải thích để mang lại độ chính xác cao hơn trong việc phân loại người học ngoại ngữ tiếng Anh, từ đó đáp ứng nhu cầu về kế hoạch học tập phù hợp với các nhóm kết quả khác nhau. Là một bài kiểm tra được

thiết kế gần đây, NGSLT có những đặc điểm được xem là khắc phục các vấn đề gây khúc mắc của hai bài kiểm tra VSL và VLT. Bài kiểm tra NGSLT bao gồm 100 câu hỏi, 20 câu hỏi được lựa chọn với mỗi dải từ để đại diện cho khoảng biến thiên và mức độ trung bình về độ khó của mỗi dải từ. Trong mỗi câu hỏi, từ đích được cung cấp và sau đó là một câu sử dụng từ trong ngữ cảnh không xác định. Tiếp theo là bốn lựa chọn câu trả lời trong đó có ba phương án nhiễu và phương án đúng. Cách diễn giải kết quả của NGSLT cung cấp thông tin về các cấp độ từ vựng như bài kiểm tra VLT, từ đó mang lại những gợi ý hữu ích cho thí sinh trong việc tích lũy từ vựng sau bài kiểm tra. Ngoài phiên bản gốc tiếng Anh, Stoeckel và cộng sự (2018) đã phát triển phiên bản song ngữ tiếng Nhật và cung cấp các bằng chứng xác trị đáng tin cậy.

Xét trong bối cảnh dạy và học ngoại ngữ Việt Nam, bài kiểm tra NGSLT có thể được xem là một lựa chọn phù hợp để đánh giá lượng từ vựng của người học, cũng như giúp xác định mục tiêu từ vựng trong kế hoạch dạy và học và lựa chọn các giải pháp dạy và học phù hợp. Ngoài bài kiểm tra về danh sách từ thông dụng, dự án NGS� cũng cung cấp các công cụ hỗ trợ như danh sách từ vựng, website học từ hay các phần mềm cài đặt trên điện thoại mang lại nhiều cơ hội và trải nghiệm khác nhau để người học có thể tích lũy vốn từ vựng thông dụng như một nhiệm vụ khả thi trong quá trình phát triển năng lực ngôn ngữ tiếng Anh và rèn luyện các kỹ năng thực hành ngôn ngữ của mình.

Với các dự án phát triển công cụ kiểm tra đánh giá từ vựng cho người học Tiếng Anh tại Việt Nam, các nhà nghiên cứu có thể xem xét các gợi ý của nhóm tác giả NGS�T như việc phát triển các phiên bản tương đương phục vụ cho việc tiến hành kiểm tra đánh giá thường xuyên và phiên bản song ngữ với các ngôn ngữ khác cũng như việc thực hiện các nghiên cứu xác trị cho bài kiểm tra NGS�T. Bên cạnh đó, đề xuất của Stoeckel và cộng sự (2021) về việc áp dụng lý thuyết ứng đáp IRT để phát triển đề kiểm tra thích ứng cũng nên được

xem xét nhằm mang lại những bài kiểm tra từ vựng với ít câu hỏi hơn, tiết kiệm thời gian hơn mà vẫn mang lại kết quả chính xác về năng lực thí sinh.

1.2.3. Khoảng trống nghiên cứu

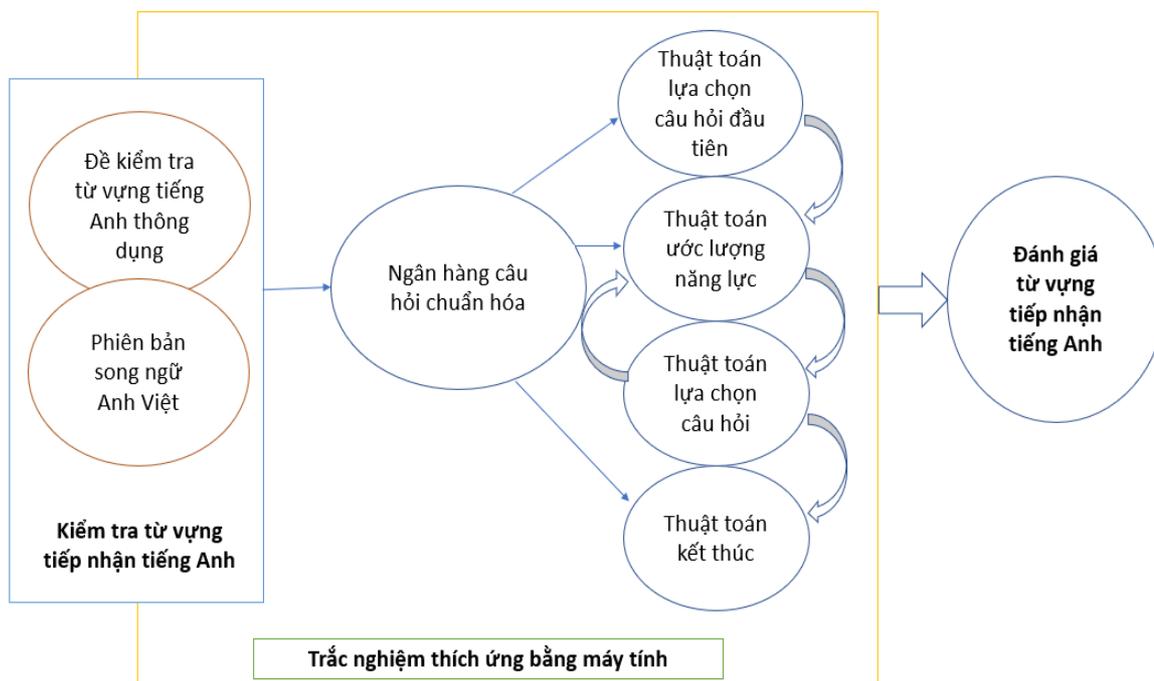
Việc tổng quan các nghiên cứu và các bài kiểm tra này đã đưa ra kết luận các bài kiểm tra từ vựng đều có giá trị thực tiễn và đang được sử dụng rộng rãi trong dạy và học tiếng Anh cũng như trong nghiên cứu. Việc lựa chọn một bài kiểm tra tùy thuộc vào mục đích sử dụng khác nhau của người học, người dạy và các nhà nghiên cứu, cũng như mức độ quen thuộc của người học với các dạng thức kiểm tra, từ đó mang lại những trải nghiệm tích cực cũng như kết quả thể hiện chính xác năng lực ngoại ngữ của thí sinh. Ngoài ra, qua việc tổng lược các bài kiểm tra từ vựng tiếp nhận tiếng Anh, một số đề xuất được đưa ra cần sự quan tâm đầu tư trong các dự án về kiểm tra từ vựng trong đó có nghiên cứu trong luận án đang thực hiện, cụ thể là các vấn đề về xác định mục đích kiểm tra rõ ràng, lấy mẫu từ kiểm tra phù hợp mục đích kiểm tra, xác định định dạng câu hỏi phù hợp, phát triển phiên bản song ngữ, áp dụng công nghệ máy tính trong thiết kế và tiến hành kiểm tra, tiến hành đánh giá hiệu quả của bài kiểm tra nghiêm túc.

Bên cạnh đó, từ việc tóm lược các vấn đề về từ vựng trong các văn bản quy định và các nghiên cứu đã được thực hiện về kiến thức từ vựng tiếng Anh của người học ngoại ngữ tiếng Anh ở Việt Nam, nhiều vấn đề cần được xem xét. Thứ nhất, danh sách từ vựng tiếng Anh được sử dụng trong kiểm tra đánh giá với người học ngoại ngữ tiếng Anh ở Việt Nam chưa được bàn đến cụ thể và đi đến thống nhất. Danh sách từ nào, khía cạnh ngôn ngữ nào và hình thức kiểm tra đánh giá nào được sử dụng để xác định được lượng từ mục tiêu vẫn chưa được xác định một cách cụ thể. Thứ hai, trong thực tế, theo kết quả của một số lượng không nhiều các nghiên cứu gần đây kiểm tra từ vựng của người

học tiếng Anh ở Việt Nam, học sinh phổ thông và sinh viên đại học có lượng từ vựng rất hạn chế, không đạt được yêu cầu về lượng từ theo quy định (Vu & Peters, 2021). Thứ ba, các công cụ kiểm tra được sử dụng cũng khác nhau, và chưa tương thích với mục tiêu đào tạo, cụ thể Nguyen và Nation (2011) dùng phiên bản song ngữ của VST của Nation và Belgar (2007); Vu và Nguyen (2019) dùng VLT của Schmitt và cộng sự (2001); Nguyen (2000) và Dang (2020) dùng VLT bản cập nhật của Webb và cộng sự (2017). Các công cụ này đã được tổng lược trong phần thảo luận trước, hướng tới đối tượng học tiếng Anh rất rộng, không hướng trực tiếp đến đối tượng người học ngoại ngữ tiếng Anh ở Việt Nam.

Từ những vấn đề này có thể đi đến nhận định rằng việc xây dựng các công cụ đánh giá kiến thức từ vựng tiếng Anh hướng tới đối tượng người học tiếng Anh ở Việt Nam là rất cần thiết. Với quy mô của luận án, công cụ được thiết kế hướng tới kiến thức cơ bản nhất là kiến thức từ vựng tiếp nhận của các đối tượng người học tiếng Anh có trình độ bậc 1-2-3 để hỗ trợ người dạy, người học cũng như những nhà nghiên cứu trong lĩnh vực dạy và học tiếng Anh tại Việt Nam. Danh sách từ vựng tiếng Anh thông dụng mới (Browne, 2013) với lượng từ là 2.801 gần với các yêu cầu về lượng từ của các trình độ 1-3. Thêm vào đó, việc xây dựng công cụ kiểm tra từ vựng có thể thừa kế điểm mạnh từ các bài kiểm tra từ vựng đã và đang được sử dụng rộng rãi trong nhiều năm qua, bao gồm định dạng câu hỏi hay phiên bản song ngữ. Danh sách từ thông dụng cần được lựa chọn phù hợp với đối tượng được kiểm tra. Ngoài ra, việc áp dụng trắc nghiệm thích ứng cũng hứa hẹn mang lại những ưu thế cho quá trình kiểm tra đánh giá. Chỉ khi có được một công cụ phù hợp hiệu quả mới có thể đánh giá được chính xác năng lực người học cũng như xác định được thực trạng vấn đề để hướng tới việc cải thiện hay nâng cao chất lượng dạy và học tiếng Anh ở Việt Nam nói chung.

1.3. Đề xuất mô hình nghiên cứu



Hình 1.10: Mô hình nghiên cứu

Sau quá trình nghiên cứu cơ sở lý luận và tổng quan nghiên cứu về trắc nghiệm thích ứng, kiểm tra từ vựng tiếp nhận tiếng Anh, luận án đi tới việc đề xuất mô hình nghiên cứu hướng tới mục đích phát triển đề trắc nghiệm thích ứng đánh giá từ vựng tiếp nhận tiếng Anh của đối tượng người học ngoại ngữ tiếng Anh tại Việt Nam. Mô hình dựa trên cơ sở lý luận từ các khung lý thuyết của các học giả uy tín trong ngành bao gồm Bachman và Palmer (1996), Read và Chapelle (2001) và Thompson và Weiss (2011) với mục đích đánh giá từ vựng tiếp nhận tiếng Anh của thí sinh. Mô hình thể hiện việc áp dụng trắc nghiệm thích ứng trong kiểm tra từ vựng tiếng Anh để thực hiện mục đích đo lường và đánh giá của luận án. Sử dụng các thuật toán được thiết lập của hệ thống trắc nghiệm thích ứng UEd-CAT bao gồm thuật toán lựa chọn câu hỏi đầu tiên, thuật toán ước lượng năng lực, thuật toán lựa chọn câu hỏi và thuật toán kết thúc bài trắc nghiệm, luận án tập trung vào việc xây dựng và chuẩn hóa ngân hàng câu hỏi đánh giá từ vựng tiếp nhận tiếng Anh, một khía cạnh cơ

bản nhất làm nền tảng cho quá trình học của người học ngoại ngữ tiếng Anh ở Việt Nam. Mô hình cũng thể hiện việc lựa chọn đề kiểm tra từ vựng tiếng Anh thông dụng NGS LT và xây dựng phiên bản song ngữ Anh Việt để làm nội dung đầu vào của ngân hàng câu hỏi trắc nghiệm thích ứng. Các vấn đề về thử nghiệm đề mẫu, quy trình phát triển phiên bản song ngữ và thử nghiệm đánh giá sẽ được trình bày trong các chương sau của luận án.

1.4. Kết chương

Chương 1 đã trình bày cơ sở lý luận về lý thuyết khảo thí hiện đại, lý luận về trắc nghiệm thích ứng bằng máy tính và lý luận về đánh giá kiến thức từ vựng tiếp nhận tiếng Anh; và tiến hành tổng quan các nghiên cứu để đi đến kết luận việc phát triển đề kiểm tra từ vựng tiếng Anh sử dụng trắc nghiệm thích ứng bằng máy tính là rất cần thiết, phù hợp xu thế giáo dục trong bối cảnh của Việt Nam và thế giới, hứa hẹn những đóng góp tích cực vào lĩnh vực kiểm tra và đào tạo ngôn ngữ mà các nhà nghiên cứu hàng đầu trong ngành đã chỉ ra.

Chương 1 cũng đã tiến đến việc đề xuất mô hình nghiên cứu hướng tới việc phát triển đề trắc nghiệm thích ứng đánh giá từ vựng tiếp nhận tiếng Anh dành cho đối tượng người học ngoại ngữ tiếng Anh tại Việt Nam. Việc sử dụng trắc nghiệm thích ứng trong kiểm tra đánh giá kiến thức từ vựng tiếp nhận tiếng Anh có thể tận dụng những lợi thế của trắc nghiệm thích ứng để khắc phục các hạn chế của các bài kiểm tra từ vựng tiếng Anh đã được phát triển, mang lại hiệu quả trải nghiệm tốt hơn cho thí sinh, hỗ trợ quá trình dạy và học ngoại ngữ tiếng Anh tại Việt Nam. Tuy nhiên, việc xây dựng hệ thống trắc nghiệm thích ứng đòi hỏi nỗ lực từ nhiều phía trong việc thiết kế, duy trì, quản trị. Việc sử dụng nguồn lực sẵn có hiện tại, hệ thống trắc nghiệm thích ứng UEd-CAT của trường ĐHGĐ – ĐHQGHN và sự hỗ trợ tích cực từ nhóm xây dựng và quản trị hệ thống có thể giảm đi những gánh nặng và thách thức với nhà nghiên cứu.

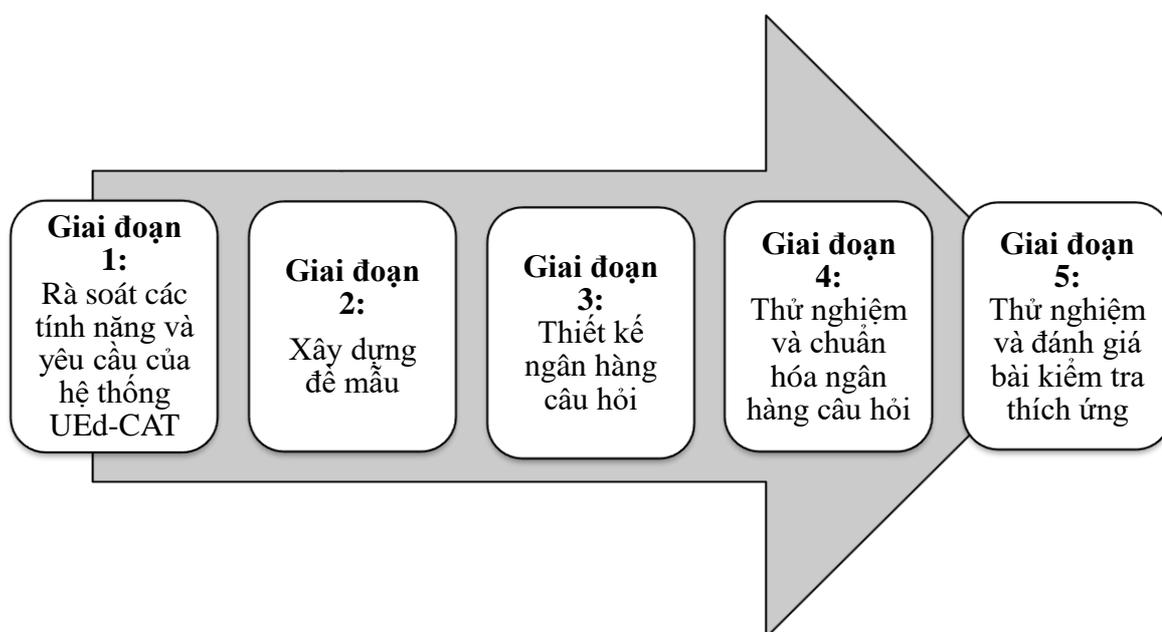
CHƯƠNG 2: PHƯƠNG PHÁP LUẬN NGHIÊN CỨU

Chương 2 được chia thành các nội dung chính: quy trình nghiên cứu, phương pháp nghiên cứu, quá trình lấy mẫu, công cụ nghiên cứu, phân tích dữ liệu và các vấn đề về đạo đức nghiên cứu. Từ các nội dung này, luận án hoàn thành việc xây dựng bảng đặc tả đề trắc nghiệm đánh giá kiến thức từ vựng tiếp nhận tiếng Anh và phân tích đề mẫu. Bảng đặc tả và kết quả phân tích đề mẫu được sử dụng để chuẩn hóa ngân hàng câu hỏi của hệ thống trắc nghiệm thích ứng và so sánh kết quả thực nghiệm của đề truyền thống và đề trắc nghiệm thích ứng.

2.1. Quy trình nghiên cứu

Quy trình nghiên cứu của luận án bao gồm các bước rà soát, thu thập dữ liệu và phân tích dữ liệu, cụ thể được chia thành năm giai đoạn (Hình 2.1). Giai đoạn 1 là giai đoạn tìm hiểu nguyên lý hoạt động của hệ thống trắc nghiệm UEd-CAT, bao gồm: yêu cầu về ngân hàng câu hỏi, thuật toán lựa chọn câu hỏi đầu tiên, thuật toán lựa chọn câu hỏi, thuật toán tính điểm, và thuật toán kết thúc. Giai đoạn 2 là giai đoạn xây dựng đề mẫu. Đề mẫu là phiên bản song ngữ Anh Việt của đề kiểm tra từ vựng tiếng Anh thông dụng NGS LT (Stoeckel và Bennett, 2015). Giai đoạn bao gồm việc biên soạn đề mẫu, chạy thử nghiệm, phân tích và điều chỉnh đề mẫu cũng như thiết lập bảng đặc tả bài trắc nghiệm song ngữ đánh giá từ vựng tiếp nhận tiếng Anh. Giai đoạn 3 là thiết kế ngân hàng câu hỏi đánh giá từ vựng tiếp nhận tiếng Anh, bao gồm các bước soạn thảo câu hỏi và xin ý kiến chuyên gia để hoàn thành bộ câu hỏi thô song ngữ với một số lượng lớn câu hỏi – mục tiêu là xây dựng ngân hàng 500 câu hỏi để sử dụng trong hệ thống trắc nghiệm thích ứng UEd-CAT theo quy trình của các nghiên cứu trước đây (Le, Tang và cộng sự, 2019; Le, Tran và cộng sự, 2019; Nguyen và cộng sự, 2021). Giai đoạn 4 thực hiện việc thử nghiệm và chuẩn hóa

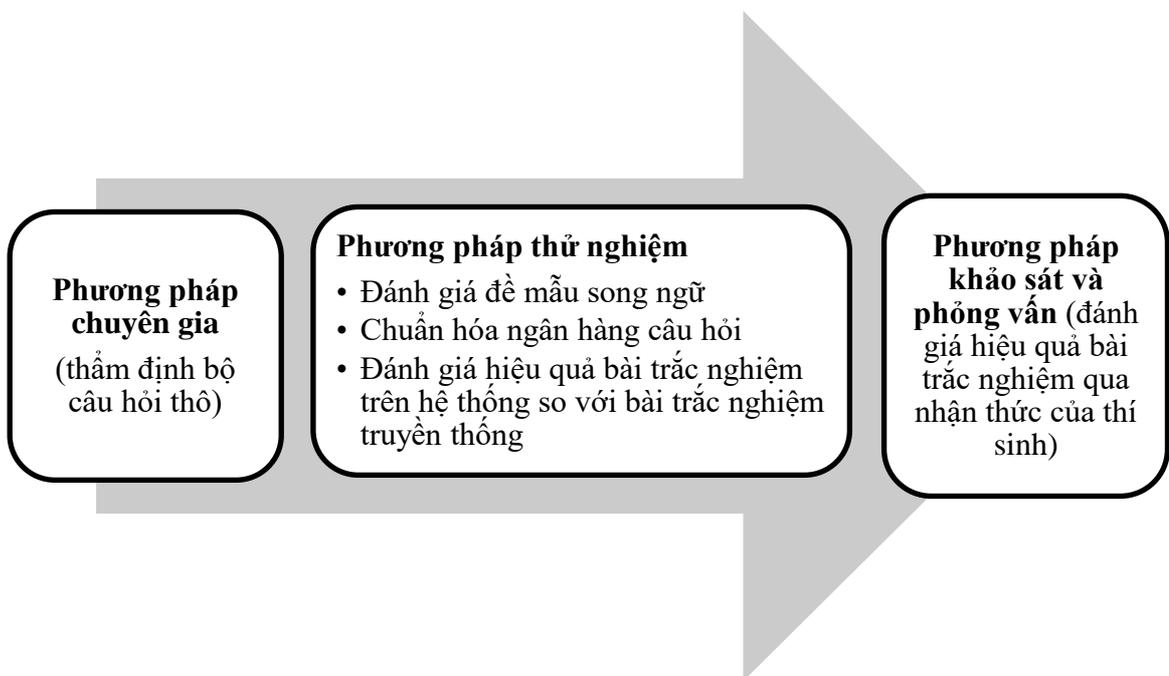
ngân hàng câu hỏi đảm bảo các yêu cầu của hệ thống trắc nghiệm thích ứng còn bao gồm việc tiến hành thử nghiệm bộ câu hỏi thô, phân tích kết quả thử nghiệm theo lý thuyết IRT, và chỉnh sửa, hoàn thiện ngân hàng câu hỏi, phục vụ trả lời câu hỏi nghiên cứu số 1. Giai đoạn 5 là giai đoạn thử nghiệm và đánh giá bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh trên hệ thống trắc nghiệm UEd-CAT, phục vụ trả lời câu hỏi nghiên cứu số 2. Việc đánh giá dựa vào kết quả thử nghiệm với thí sinh để xác định tính chính xác, đồng thời so sánh hiệu quả của đề trắc nghiệm thích ứng với đề trắc nghiệm truyền thống theo quy trình của các nghiên cứu đã thực hiện trước đây trên cùng hệ thống (Le & Nguyen, 2021; Nguyen và cộng sự, 2021). Trong giai đoạn này, điều tra khảo sát và phỏng vấn được thực hiện để tìm hiểu quan điểm của thí sinh về các trải nghiệm với bài kiểm tra thích ứng đánh giá từ vựng tiếp nhận tiếng Anh, cũng như những đề xuất để hoàn thiện bài kiểm tra cũng như hệ thống nói chung. Từ kết quả nghiên cứu, tác giả đưa ra những khuyến nghị về việc duy trì và phát triển bài kiểm tra thích ứng trên máy tính.



Hình 2.1: Quy trình nghiên cứu

2.2. Phương pháp nghiên cứu

Phương pháp nghiên cứu kết hợp (mixed method approach) được lựa chọn sử dụng trong luận án với việc sử dụng kết hợp phương pháp nghiên cứu định lượng và định tính trong các bước thu thập và phân tích dữ liệu nghiên cứu. Dữ liệu định lượng được thu thập bao gồm dữ liệu từ thử nghiệm và dữ liệu từ khảo sát với thí sinh. Dữ liệu định tính bao gồm dữ liệu từ phương pháp chuyên gia và dữ liệu từ phỏng vấn bán cấu trúc với thí sinh. Việc sử dụng kết hợp phương pháp định tính và định lượng trong luận án được thực hiện với mục đích phát huy được điểm mạnh của các cách tiếp cận nghiên cứu đơn lẻ, từ đó thu nhận được các kết quả đa diện, giúp làm rõ vấn đề nghiên cứu từ nhiều góc độ khác nhau (Creswell & Plano Clark, 2007). Hình 2.2 sau đây mô tả mục đích sử dụng các phương pháp nghiên cứu trong luận án.



Hình 2.2: Mục đích sử dụng các phương pháp nghiên cứu

2.2.1. Phương pháp nghiên cứu định lượng

2.2.1.1. Phương pháp thử nghiệm

Thiết kế nghiên cứu thử nghiệm hay thực nghiệm (experimental research) được sử dụng phổ biến trong các nghiên cứu giáo dục, đặc biệt là trong các nghiên cứu thiết kế, xây dựng hay phát triển các sáng kiến giáo dục hay các công cụ đo lường đánh giá mới. Phương pháp thử nghiệm cũng được đề xuất trong khung thiết kế trắc nghiệm thích ứng của Thompson và Weiss (2011). Tất cả các nghiên cứu trước đây được thực hiện với hệ thống trắc nghiệm thích ứng UEd-CAT đều báo cáo kết quả sau quá trình thử nghiệm với đối tượng thí sinh hướng tới (Le & Nguyen, 2021; Le, Tang và cộng sự, 2019; Le, Tran và cộng sự, 2019; Nguyen và cộng sự, 2021).

Trong nghiên cứu này, thử nghiệm được tiến hành trong ba đợt: đợt 1 là thử nghiệm để đánh giá đề mẫu; đợt 2 để chuẩn hóa ngân hàng câu hỏi, và đợt 3 là thử nghiệm để đánh giá bài trắc nghiệm thích ứng. Đợt 1 tiến hành với người học ngoại ngữ Tiếng Anh có trình độ bậc 1-2-3 từ đó đưa ra những đánh giá và nhận định với đề mẫu song ngữ được phát triển từ đề NGSLT đơn ngữ tiếng Anh. Với đợt 2, việc tiến hành thử nghiệm là để chọn lọc các câu hỏi phù hợp với mô hình và mục đích đánh giá cũng như ước lượng các tham số của các câu hỏi trong ngân hàng câu hỏi. Bộ câu hỏi thô được chia thành các đề tuân theo bảng đặc tả của đề mẫu và có thiết kế câu hỏi neo để thực hiện bước cân bằng sau đó. Sau quá trình phân tích các đề riêng lẻ, nhà nghiên cứu tiến hành cân bằng đề để định cỡ các tham số của tất cả các câu hỏi trong ngân hàng câu hỏi trên một thang đo duy nhất, đảm bảo yêu cầu chuẩn hóa của ngân hàng câu hỏi trắc nghiệm thích ứng. Từ việc thử nghiệm bộ câu hỏi với đối tượng thí sinh tham gia kiểm tra, và cân bằng, ngân hàng câu hỏi chuẩn hóa, chỉnh sửa và hoàn thiện. Với đợt 3, thử nghiệm được tiến hành để đánh giá hiệu quả của

bài kiểm tra trắc nghiệm thích ứng thông qua các lượt làm bài kiểm tra của các thí sinh, tương quan kết quả làm bài kiểm tra thích ứng với kết quả làm bài kiểm tra cố định. Đối tượng tham gia thử nghiệm đợt 3 được mời thực hiện một đề hoàn chỉnh 100 câu trước khi thực hiện bài trắc nghiệm thích ứng trên hệ thống. Kết quả thu được sau đó sẽ được phân tích để so sánh kết quả đánh giá của bài trắc nghiệm thích ứng với bài trắc nghiệm cố định.

2.2.1.2. Phương pháp điều tra khảo sát

Phương pháp nghiên cứu điều tra khảo sát là một công cụ phổ biến và hữu hiệu để thu thập dữ liệu về các khía cạnh khác nhau của hành vi, thái độ và quan điểm của con người (Fowler Jr, 2014). Phương pháp này được sử dụng rộng rãi trong khoa học xã hội, nghiên cứu thị trường, thăm dò dư luận và các lĩnh vực khác để giải quyết các câu hỏi nghiên cứu và đưa ra quyết định dựa trên dữ liệu. Điều tra khảo sát, khi được tiến hành một cách chặt chẽ và có đạo đức, sẽ đóng góp đáng kể vào sự hiểu biết về thế giới và cung cấp thông tin cho quá trình ra quyết định trong nhiều lĩnh vực khác nhau.

Trong nghiên cứu này, điều tra khảo sát được tiến hành sau đợt thử nghiệm với hệ thống trắc nghiệm thích ứng, tìm hiểu về ý kiến của thí sinh với các tiêu chí của trắc nghiệm thích ứng. Nghiên cứu tin tưởng rằng việc chia sẻ về trải nghiệm của thí sinh với bài kiểm tra thích ứng sẽ mang lại những thông tin hữu ích cho việc đánh giá của chính thí sinh về trải nghiệm của họ với bài kiểm tra thích ứng bên cạnh những phân tích về kết quả thực hiện của thí sinh.

2.2.1.3. Phương pháp phân tích dữ liệu định lượng

Phân tích và cân bằng đề thử nghiệm

Nghiên cứu đang thực hiện áp dụng mô hình một tham số, hay mô hình Rasch để hiệu chuẩn các câu hỏi thô. Phương trình của mô hình Rasch (Carlson, 2020) như sau:

$$P(u_i = 1|\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

Trong đó, θ là ước lượng năng lực thí sinh, P là khả năng trả lời đúng câu hỏi i của thí sinh với mức năng lực θ , u_i là câu trả lời của thí sinh cho câu hỏi, b_i là độ khó của câu hỏi i . Các tham số về độ khó của câu hỏi và ước tính năng lực thí sinh có cùng thang đo, thường nằm trong khoảng từ -3 đến +3.

Nghiên cứu này lựa chọn sử dụng phần mềm Conquest 2.0, áp dụng mô hình Rasch để tiến hành phân tích chất lượng đề và câu hỏi, đồng thời xác định các tham số câu hỏi của các đề thử nghiệm.

Sau khi phân tích chất lượng đề và câu hỏi của các đề thử nghiệm, các câu hỏi không phù hợp với mô hình sẽ bị loại bỏ, nhà nghiên cứu tiến hành cân bằng đề với phần mềm R, gói equateIRT, để đưa các tham số câu hỏi về cùng một thang đo của ngân hàng câu hỏi trắc nghiệm thích ứng. Cụ thể, phương pháp tiến hành cân bằng độ khó của các câu hỏi được thực hiện trong các nghiên cứu này là phương pháp “câu hỏi neo” (David & Ida, 2019). Trong phương pháp cân bằng độ khó của câu hỏi thi bằng câu hỏi neo, các tham số độ khó (b), độ phân biệt (a), mức năng lực (θ) của câu hỏi trong đề thi $s+1$ được tính thông qua đề thi s bằng công thức sau:

$$\theta_{s+1} = A_{s,s+1}\theta_s + B_{s,s+1}, (1)$$

$$a_{s+1} = \frac{a_s}{A_{s,s+1}}, \quad (2)$$

$$b_{s+1} = A_{s,s+1} b_s + B_{s,s+1} \quad (3)$$

Trong đó $n_{s,s+1}$ là các câu hỏi neo giữa đề thi s và đề thi $s+1$; $A_{s,s+1}$, $B_{s,s+1}$ là các hệ số cân bằng độ khó giữa hai đề thi. Trong nghiên cứu này, để tính các hệ số cân bằng chúng tôi sử dụng phương pháp “mean – mean” (Lloyd & Hoover, 1980), theo đó, hệ số cân bằng được tính như sau:

$$A_{s,s+1} = \left(\prod_{j=1}^{n_{s,s+1}} \frac{a_{sj}}{a_{(s+1)j}} \right)^{\frac{1}{n_{s,s+1}}} \quad (4)$$

$$B_{s,s+1} = \frac{1}{n_{s,s+1}} \sum_{j=1}^{n_{s,s+1}} b_{(s+1)j} - A_{s,s+1} \frac{1}{n_{s,s+1}} \sum_{j=1}^{n_{s,s+1}} b_{sj} \quad (5)$$

Phân tích kết quả thử nghiệm đề trắc nghiệm thích ứng

Với ngân hàng câu hỏi được chuẩn hóa, nhà nghiên cứu tiến hành thử nghiệm và thu thập dữ liệu thử nghiệm trên hệ thống trắc nghiệm thích ứng. Các bước phân tích dữ liệu thử nghiệm được thực hiện để cung cấp các bằng chứng xác trị đề trắc nghiệm thích ứng. Các bước phân tích bao gồm việc sử dụng SPSS để tiến hành thống kê mô tả, tương quan giữa số câu trả lời đúng và năng lực thí sinh, tương quan giữa điểm năng lực thí sinh trên hệ thống trắc nghiệm thích ứng và điểm thi với bài kiểm tra truyền thống. Dựa trên các kết quả phân tích, nhà nghiên cứu cung cấp các bằng chứng về độ tin cậy của bài trắc nghiệm thích ứng trên máy tính để đánh giá từ vựng tiếp nhận tiếng Anh của thí sinh.

2.2.2. Phương pháp nghiên cứu định tính

2.2.2.1. Phương pháp chuyên gia

Phương pháp chuyên gia (tiếng Anh là expert method hay expert opinion method) là một phương pháp được sử dụng rộng rãi không chỉ trong khoa học xã hội và tâm lý học mà còn trong khoa học giáo dục và được đánh giá là một trong những phương pháp thích hợp nhất để thu thập, phân tích và đánh giá thông tin (Iriste & Katane, 2018).

Phương pháp chuyên gia được định nghĩa là phương pháp nghiên cứu nghiêm túc và có hệ thống được sử dụng để phân tích các vấn đề phức tạp, dựa trên sự tham khảo ý kiến của các chuyên gia, nhằm đạt được sự đồng thuận đáng tin cậy nhất (Bayona-Ore và cộng sự, 2018). Theo thống kê của Protasevich (2014, trích trong Iriste & Katane, 2018) tần suất của phương pháp chuyên gia là 38,4% trong tổng số các phương pháp nghiên cứu thực nghiệm của các nghiên cứu sinh Nga, chuyên ngành sư phạm. Tuy chưa có một nghiên cứu thống kê nào tương tự tại Việt Nam nhưng không thể phủ nhận vị trí cũng như mức độ sử dụng phổ biến của phương pháp này trong các nghiên cứu giáo dục tại Việt Nam.

Luận án cũng sử dụng phương pháp chuyên gia để thẩm định bộ câu hỏi thô trước khi đưa bộ câu hỏi thô tiến hành thử nghiệm với thí sinh. Do việc thiết kế bộ câu hỏi thô không được thực hiện bởi một nhóm, mà bởi cá nhân nghiên cứu sinh nên việc sử dụng phương pháp chuyên gia giúp đảm bảo chất lượng bộ câu hỏi thô trước khi đưa vào thử nghiệm với lượng thí sinh lớn.

Sau khi bộ câu hỏi thô được thiết kế hoàn chỉnh, nhà nghiên cứu chuyển bộ câu hỏi đến các giảng viên tiếng Anh có kinh nghiệm giảng dạy tiếng Anh nhiều năm cũng như kinh nghiệm viết câu hỏi thi. Nhóm chuyên gia làm việc

độc lập với bộ câu hỏi thô cùng với các tiêu chí thẩm định do nhà nghiên cứu cung cấp trong khoảng thời gian từ 2-4 tuần. Sau khi thu thập lại ý kiến từ nhóm chuyên gia, nhà nghiên cứu tiến hành chỉnh sửa dựa vào các ý kiến của nhóm chuyên gia trước khi đưa bộ câu hỏi thô vào thử nghiệm.

2.2.2.2. Phương pháp phỏng vấn

Bên cạnh phương pháp điều tra khảo sát, nghiên cứu sử dụng phương pháp phỏng vấn bán cấu trúc để tìm hiểu về quan điểm của thí sinh – những người có trải nghiệm trực tiếp – với bài kiểm tra thích ứng được thiết kế.

Phương pháp nghiên cứu phỏng vấn đóng một vai trò quan trọng trong nghiên cứu định tính, cho phép các nhà nghiên cứu khám phá các hiện tượng phức tạp, đi sâu vào quan điểm của các cá nhân và hiểu sâu hơn về kinh nghiệm và quan điểm của họ (Seidman, 2019). Các cuộc phỏng vấn cung cấp một nguồn dữ liệu phong phú, cho phép các nhà nghiên cứu nắm bắt được các sắc thái và chi tiết theo ngữ cảnh mà có thể không dễ dàng có được thông qua các phương pháp nghiên cứu khác, cụ thể là phương pháp điều tra khảo sát trong nghiên cứu này.

Phỏng vấn bán cấu trúc được lựa chọn sử dụng trong nghiên cứu này. Nhà nghiên cứu dựa vào một bảng câu hỏi được xác định trước cho tất cả những người tham gia để đảm bảo tính nhất quán trong các cuộc phỏng vấn và tạo điều kiện thuận lợi cho việc hệ thống và phân tích dữ liệu sau này. Các cuộc phỏng vấn đều được thực hiện trong khoảng thời gian từ 30-45 phút và không diễn ra cứng nhắc mà để khoảng mở cho người tham gia có thể linh hoạt và thoải mái bày tỏ suy nghĩ và kinh nghiệm của mình theo chỉ dẫn của các nhà nghiên cứu đi trước như Creswell (2013) và Seidman (2019).

2.2.2.3. Phương pháp phân tích nội dung định tính

Phương pháp phân tích nội dung được sử dụng trong việc phân tích kết quả từ phương pháp chuyên gia thẩm định bộ câu hỏi thô và phương pháp phỏng vấn thí sinh về trải nghiệm với đề trắc nghiệm thích ứng. Các kết quả định tính sẽ được phân tích và báo cáo theo nhóm. Với phương pháp chuyên gia, kết quả thẩm định bao gồm đánh giá chất lượng chung của bộ câu hỏi và các đề xuất trình sửa về nội dung và ngôn ngữ, sau đó được diễn giải để định hướng cho các chỉnh sửa cần thiết với bộ câu hỏi thô trước khi bước thử nghiệm được tiến hành. Với phương pháp phỏng vấn, các nội dung phỏng vấn được chuyển sang dạng văn bản, sau đó phân tích và tổng hợp theo nhóm phản hồi, từ đó chỉ ra các ý kiến cũng như diễn giải của thí sinh, bổ sung kết quả định lượng từ khảo sát về nhận thức của thí sinh với bài trắc nghiệm thích ứng, nhờ đó cung cấp thêm bằng chứng đánh giá bài trắc nghiệm thích ứng thông qua nhận thức của thí sinh hướng tới những nhận định tổng thể, hữu ích cho việc đánh giá và phát triển hệ thống, cũng như đóng góp vào lĩnh vực nghiên cứu còn tương đối mới mẻ trong đo lường và đánh giá ngôn ngữ tại Việt Nam.

2.3. Quá trình lấy mẫu

2.3.1. Mẫu của phương pháp chuyên gia

Với mục đích thẩm định bộ câu hỏi thô, nghiên cứu lựa chọn người tham gia phương pháp chuyên gia bằng phương pháp lấy mẫu có chủ đích (purposive sampling) với 2 tiêu chí chung: (1) có hiểu biết và trải nghiệm với việc dạy/kiểm tra tiếng Anh, cụ thể là từ vựng tiếng Anh, (2) tình nguyện và sẵn sàng tham gia nghiên cứu. Có bảy chuyên gia đã nhận lời mời từ nhà nghiên cứu và được cung cấp thông tin về nghiên cứu cũng như mục đích đạt được từ phương pháp chuyên gia. Tất cả là thạc sĩ chuyên ngành giảng dạy tiếng Anh và một số có chứng chỉ khóa học ra đề hay thiết kế đề chuẩn hóa. Thông tin của

nhóm chuyên gia được tóm lược trong bảng 2.1. Số lượng chuyên gia tham gia tuy không lớn nhưng vẫn đảm bảo tính khách quan và hỗ trợ nhà nghiên cứu đảm bảo chất lượng bộ câu hỏi thô.

Bảng 2.1: Thông tin mẫu của phương pháp chuyên gia

STT	Tên (đã thay đổi)	Giới tính	Tuổi	Trình độ	Chứng chỉ ra đề/thiết kế đề chuẩn hóa	Số năm kinh nghiệm giảng dạy tiếng Anh
1	Thục	Nữ	50	Thạc sĩ	Có	28
2	Hà	Nữ	47	Thạc sĩ	Có	25
3	Ngân	Nữ	46	Thạc sĩ	Không	24
4	Huy	Nam	42	Thạc sĩ	Không	20
5	Chuyên	Nữ	40	Thạc sĩ	Có	18
6	Khánh	Nữ	38	Thạc sĩ	Có	15
7	Hoàng	Nam	34	Thạc sĩ	Có	11

2.3.2. Mẫu tham gia thử nghiệm

Các đợt thử nghiệm trong luận án được thực hiện tại một trường đại học tại Hà Nội. Cần lưu ý thêm rằng, mẫu nghiên cứu trong luận án là sinh viên năm nhất của trường theo học tại các chuyên ngành kỹ thuật với trình độ tiếng Anh theo kết quả kiểm tra đầu vào là các bậc 1-2-3. Quá trình lấy mẫu được thực hiện theo phương pháp lấy mẫu thuận tiện ngay sau thời gian nhập học của sinh viên, với sự đồng thuận và tự nguyện tham gia của các em sinh viên trong cả ba đợt thử nghiệm.

Đợt thử nghiệm đầu tiên thực hiện với đề mẫu – đề trắc nghiệm từ vựng tiếp nhận tiếng Anh phiên bản song ngữ Anh-Việt. Thành phần tham gia bao gồm 225 sinh viên. Quy mô mẫu thử nghiệm gồm hơn 150 thí sinh với mỗi đề

đáp ứng các yêu cầu do Şahin và Weiss (2015) khuyến nghị để đảm bảo ước tính chính xác năng lực của thí sinh và tham số câu hỏi.

Tất cả đối tượng đều tình nguyện tham gia thử nghiệm bằng cách làm bài kiểm tra từ vựng 100 câu và có thể xem kết quả số câu trả lời đúng ngay sau khi hoàn thành. Trước khi tiến hành kiểm tra, các hướng dẫn kiểm tra đã được cung cấp cho các sinh viên tham gia, sinh viên có thể hỏi bất kỳ câu hỏi liên quan trước khi bắt đầu kiểm tra. Thí sinh cần xác nhận tham gia tự nguyện và tập trung hoàn thành bài kiểm tra.

Đợt thử nghiệm thứ hai diễn ra trong một tháng với 1619 sinh viên tham gia. Sau khi tiến hành làm sạch dữ liệu, nhà nghiên cứu lấy kết quả và tiến hành phân tích để đánh giá chất lượng bộ câu hỏi thô cũng như định cỡ các tham số của câu hỏi trên cùng thang đo để đưa vào ngân hàng câu hỏi hiệu chuẩn. Bảng 2.2 cung cấp số liệu tham gia thử nghiệm 7 đề được xây dựng từ bộ câu hỏi thô, tuân theo đặc tả được xác định từ trước như sau:

Bảng 2.2: Cỡ mẫu thử nghiệm theo đề

Đề	1	2	3	4	5	6	7	Tổng
Cỡ mẫu	237	218	229	214	245	221	255	1619

Đợt thử nghiệm thứ ba được tiến hành với cỡ mẫu nhỏ hơn. Sinh viên được mời thực hiện một đề hoàn chỉnh 100 câu hỏi để xác định năng lực thí sinh theo cách thức kiểm tra truyền thống, có 290 sinh viên tham gia thực hiện. Số sinh viên này được gửi thư mời tham gia tiếp giai đoạn thử nghiệm trên hệ thống trắc nghiệm thích ứng cùng thông tin tài khoản và hướng dẫn cụ thể.

Sau đây là hướng dẫn thực hiện:

→ Link thử nghiệm hệ thống: <http://cat2.education.vnu.edu.vn/exam>

→ Chọn **Đề kiểm tra từ vựng tiếng Anh thông dụng**



→ Nhấn

Bắt đầu làm bài

→ **Đăng nhập** theo thông tin sau:

Tài khoản: tqhoan20225320@hust.cat.vn

Hình 2.3: Hướng dẫn làm bài trên hệ thống

Đợt thử nghiệm trên hệ thống trắc nghiệm thích ứng có được sự tham gia của 98 sinh viên với 209 lượt làm bài, cụ thể:

Bảng 2.3: Tổng hợp số liệu sinh viên thử nghiệm trên hệ thống

STT	Số sinh viên	Số lượt làm bài trên hệ thống
1	40	1
2	26	2
3	20	3
4	8	4
5	3	6
6	1	7
Tổng	98	209

2.3.3. Mẫu tham gia khảo sát và phỏng vấn

Sau quá trình thử nghiệm, số sinh viên có tham gia thử nghiệm trên hệ thống được mời thực hiện khảo sát ngắn về ý kiến với các tính năng của bài kiểm tra thích ứng đã thực hiện. Trong số 98 sinh viên được gửi thư mời, có 74 sinh viên, tỉ lệ 75,5%.

Bảng 2.4: Thông tin của người tham gia khảo sát

Thông tin		Tần suất	Tỉ lệ (%)
Giới tính	Nam	43	58.1
	Nữ	31	41.9
Ước lượng từ vựng	< 1500 từ	39	52.7
	1500 - 2750 từ	23	31.1
	2750 - 3250 từ	7	9.45
	3250 - 3750 từ	3	4.05
	3750 - 4500 từ	1	1.35
	> 4500	1	1.35
Tần suất trải nghiệm trắc nghiệm thích ứng	Không bao giờ	18	24.3
	Hiếm khi	32	43.25
	Thỉnh thoảng	20	27.05
	Thường xuyên	4	5.4
Thiết bị sử dụng để trải nghiệm trắc nghiệm thích ứng	Máy tính xách tay	29	39.2
	Máy tính bảng	4	5.4
	Điện thoại thông minh	40	54.05
	Máy tính để bàn	1	1.35

Bảng 2.4 trình bày thông tin cơ bản của người trả lời liên quan đến giới tính, ước lượng từ vựng, tần suất trải nghiệm với CAT và các thiết bị được sử dụng để trải nghiệm CAT. Từ bảng này có thể thấy trong số 74 người trả lời có 43 nam và 31 nữ. Hầu hết (83,78%) tự đánh giá vốn từ vựng của mình theo hai nhóm là dưới 1500 từ và 1500 - 2750 từ và cho biết tần suất trải nghiệm với CAT là khá hạn chế. Điện thoại thông minh là thiết bị được sử dụng phổ biến nhất trong trải nghiệm CAT (54,05%), tiếp theo là máy tính

xách tay (39,2%). Những số liệu này sẽ được sử dụng thêm để phân tích về nhận thức của thí sinh trong phần sau.

Bảng 2.5: Thông tin của người tham gia phỏng vấn

STT	Tên (đã thay đổi)	Chuyên ngành	Giới tính	Điểm bài trắc nghiệm thích ứng
1	NHẬT	Kinh tế và quản lý	Nam	99.85
2	NHI	Hóa và khoa học sự sống	Nữ	91.5
3	TÂM	Kinh tế và quản lý	Nữ	99.98
4	TÚ	Toán ứng dụng và tin học	Nam	83.1
5	NGA	Hóa và khoa học sự sống	Nữ	99.98
6	LAN	Hóa và khoa học sự sống	Nữ	59.6
7	THU	Hóa và khoa học sự sống	Nữ	99.96
8	BÁCH	Công nghệ thông tin và truyền thông	Nam	79.87
9	CẨM	Toán ứng dụng và tin học	Nữ	99.92
10	THANH	Cơ khí	Nữ	82.04

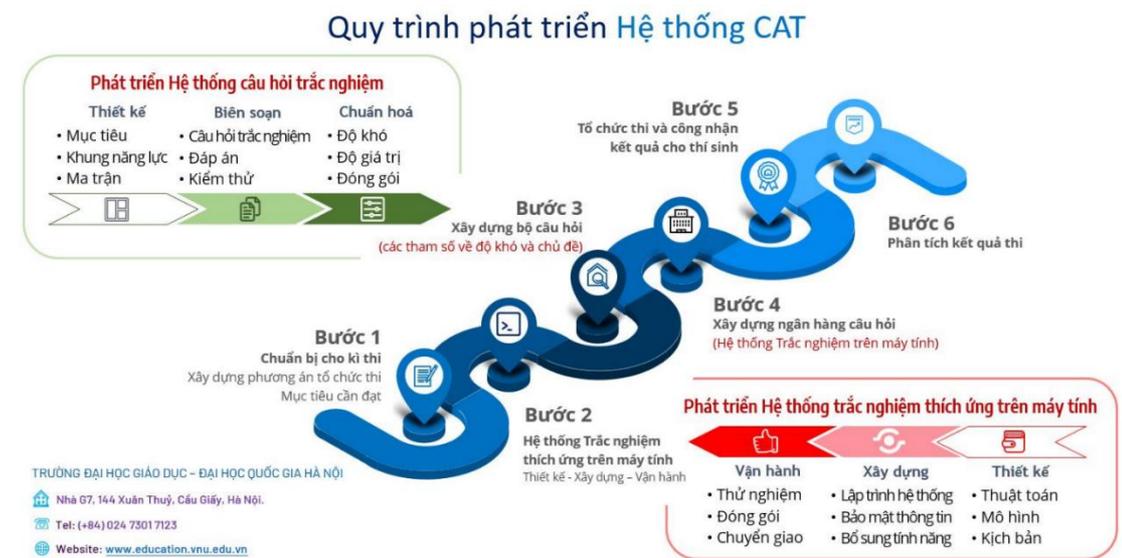
Sau quá trình khảo sát, 10 sinh viên đã nhận lời mời tham gia phỏng vấn với thông tin được trình bày trong Bảng 2.5. Để có thể thu nhận được ý kiến đa chiều, mẫu tham gia phỏng vấn được lựa chọn theo phương pháp lấy mẫu có chủ đích với nỗ lực đảm bảo tính đa dạng từ chuyên ngành đào tạo đến điểm của bài trắc nghiệm thích ứng mà sinh viên đã thực hiện trước đó.

2.4. Công cụ nghiên cứu

2.4.1. Hệ thống UEd-CAT

2.4.1.1. Thông tin chung

Hệ thống trắc nghiệm thích ứng phiên bản UEd-CAT là sản phẩm khoa học của nhóm nghiên cứu Khoa Quản trị Chất lượng – Trường ĐHQGĐ – ĐHQGHN sau một quá trình phát triển với những nỗ lực nghiên cứu chính xác, bền bỉ và hiệu quả. Cũng như các hệ thống trắc nghiệm thích ứng khác, quá trình xây dựng hệ thống UEd-CAT gồm nhiều bước từ bước chuẩn bị cho đến phân tích kết quả (Hình 2.4).



Hình 2.4: Quy trình phát triển hệ thống trắc nghiệm thích ứng (Lê Thái Hưng & Nguyễn Thái Hà, 2020)

Hệ thống UEd-CAT hiện vẫn đang được nhóm nghiên cứu tiếp tục phát triển và hoàn thiện. Cùng với sự tham gia của các giáo viên trường THPT Khoa học giáo dục, nhóm nghiên cứu đã xây dựng thành công và báo cáo những kết quả tích cực đáng ghi nhận với đề thi trắc nghiệm thích ứng đánh giá năng lực

toán học và năng lực đọc hiểu của học sinh lớp 10 với 1000 câu hỏi đã được chuẩn hoá theo lý thuyết IRT (Lê Thái Hưng và cộng sự, 2019).

2.4.1.2. Nguyên lý hoạt động

Hệ thống được phát triển dựa trên phương pháp ước lượng hậu nghiệm cực đại (Maximum a posteriori), thuật toán Gradient Descent, lý thuyết IRT mô hình một tham số và lập trình bằng ngôn ngữ PHP, Javascript. Các thuật toán này đã được tổng hợp trong các xuất bản trước đây về hệ thống (Nguyen & Le, 2018; Le & Nguyen, 2021).

Hàm biến cố hợp lý cực đại và ước lượng hậu nghiệm cực đại

Hàm biến cố hợp lý cực đại ứng với $k-1$ câu hỏi được cho bởi:

$$L(\theta / X_1, \dots, X_{k-1}) = \prod_{i=1}^{k-1} P_i(\theta)^{x_i} (1 - Q_i(\theta))^{1-x_i}$$

P_i là xác suất để trả lời đúng câu hỏi $i^{(th)}$ và được tính bằng công thức sau theo lý thuyết IRT:

$$P_i(\theta) = P(X_i = 1 / \theta, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{\exp(a_i(\theta - b_i))}{1 + \exp(a_i(\theta - b_i))}$$

Các ước lượng Bayes xem các tham số cần ước lượng là một biến ngẫu nhiên chứ không phải một hằng số. Tham số cần ước lượng sẽ được gán với một phân bố ban đầu. Ước lượng kiểu Bayes có thể sử dụng khi mà dữ liệu ít và sau đó ước lượng sẽ được cải thiện tốt hơn nếu có thêm dữ liệu. Trong suy diễn Bayes, ban đầu ta giả sử rằng θ tuân theo một phân bố gọi là phân bố tiên nghiệm (prior distribution) $f(\theta)$. Sau đó dựa vào phân bố tiên nghiệm và hàm hợp lý cực đại ta suy ra được phân bố hậu nghiệm của tham số θ :

$$g(\theta / X_1, \dots, X_{k-1}) = \frac{f(\theta)L(\theta / X_1, \dots, X_{k-1})}{\int f(\theta)L(\theta / X_1, \dots, X_{k-1})d\theta}$$

Trong suy diễn Bayes, ước lượng phân phối hậu nghiệm cực đại (Maximum a Posteriori Estimator- MAP) được giới thiệu trong IRT trong cuốn sách của Lord (1986) và được cho bởi công thức sau:

$$\hat{\theta}_{MAP} = \operatorname{argmax}_{\theta} g(\theta / X_1, \dots, X_{k-1})$$

Một loại công cụ ước lượng khác là ước lượng kỳ vọng (Expected a posteriori – EAP)

$$\hat{\theta}_{EAP} = \int_{\theta} \theta g(\theta / X_1, \dots, X_{k-1})$$

Để tìm $\hat{\theta}$ ta chuyển bài toán từ tìm cực đại của hàm số

$$\ln(f(\theta)L(\theta / X_1, \dots, X_{k-1}))$$

Nếu ta chọn phân bố tiên nghiệm của θ là phân bố chuẩn với trung bình 0, độ lệch chuẩn 1 thì:

$$f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}$$

Khi đó ta có:

$$\begin{aligned} l(\theta) &= \ln(f(\theta)L(\theta / X_1, \dots, X_{k-1})) \\ &= C - \frac{\theta^2}{2} + \sum_{i=1}^{k-1} X_i \ln \ln(P_i(\theta)) + (1 - P_i(\theta)) \end{aligned}$$

Ở đó C là hằng số. Để tìm maximum của $l(\theta)$, ta dùng thuật toán “Gradient Descent” được mô tả dưới đây.

Thuật toán Gradient Descent

Thuật toán Gradient Descent là thuật toán hiệu quả được dùng để tìm các điểm cực trị của hàm số khi mà giải phương trình đạo hàm bằng 0 khá phức tạp (Vũ Hữu Tiệp, 2018). Giả sử ta muốn tìm cực tiểu của hàm một biến $f(x)$. Ta có thể mô tả sơ lược thuật toán như sau: từ một điểm bất kỳ trên đồ thị x , ta cố gắng di chuyển điểm x về điểm mà tại đó $f(x)$ đạt giá trị cực tiểu, ký hiệu là x^* .

Điểm x sẽ di chuyển theo hướng ngược với dấu của đạo hàm trong trường hợp ta muốn tìm cực tiểu. Vì giả sử $f'(x) > 0$, thì x nằm về phía bên phải so với x^* , do đó x phải giảm để tiến tới x^* . Ngược lại nếu $f'(x) < 0$, thì x nằm về phía bên trái so với x^* , do đó x phải tăng để tiến tới x^* . Thì công thức cập nhật điểm x như sau:

$$x_k = x_{k-1} - \gamma f'(x_{k-1})$$

Ở đó tại bước đầu tiên $x_0 = x$, γ là tốc độ học (learning rate). Sau một số hữu hạn bước, điểm x sẽ di chuyển về gần điểm x^* . Theo thuật toán trên ta phải tính $f'(\theta)$ tại mỗi bước của thuật toán.

$$\begin{aligned} f'(\theta) &= \sum_{i=1}^{k-1} \frac{P_i'(\theta)}{p_i(\theta)} - (1 - X_i) \frac{P_i'(\theta)}{1 - P_i(\theta)} \\ &= \sum_{i=1}^{k-1} X_i P_i'(\theta) \left(\frac{1}{p_i} + \frac{1}{Q_i} \right) - \frac{P_i'(\theta)}{Q_i(\theta)} = \sum_{i=1}^{k-1} X_i - P_i(\theta) \end{aligned}$$

Ta mô tả thuật toán như sau:

Bước 1: Chọn điểm bắt đầu $\theta = 0$ và tốc độ học γ .

Bước 2: Cập nhật

$$\theta := \theta + \gamma \sum_{i=1}^{k-1} \left(X_i - \frac{\exp(\theta - b_i)}{1 + \exp(\theta - b_i)} \right)$$

Thuật toán tìm kiếm nhị phân (Binary search) để tìm câu hỏi tiếp theo

Tiêu chí chọn câu hỏi. Sau khi thí sinh trả lời câu hỏi $k-1$ thì năng lực tạm thời của thí sinh được ước lượng và kí hiệu là $\hat{\theta}_{k-1}$. Tiếp theo ta phải tìm câu hỏi thứ k phù hợp với mức năng lực này bằng phương pháp lựa chọn câu hỏi theo tiêu chuẩn thông tin tối đa (Maximum Information Criterion) (Van der Linden and Glas (2010)):

$$i_k = \operatorname{argmax}_j \{ I_{X_j}(\hat{\theta}_{k-1}), j \in R_k \}$$

Ở đó $I(\theta)$ là hàm thông tin Fisher:

$$I(\theta) = \frac{(P'(\theta))^2}{P(\theta)(1-P(\theta))}$$

Khi θ cố định, hàm thông tin Fisher đạt giá trị cực đại tại điểm $b = \theta$. Vì vậy câu hỏi ik được chọn là câu hỏi có độ khó gần với $\hat{\theta}_{k-1}$. Hàm thông tin đạt giá trị maximum khi: Giá trị độ khó b gần bằng giá trị năng lực θ và độ phân biệt a càng lớn. Câu hỏi thứ k được chọn là câu hỏi có độ khó gần bằng năng lực ước lượng $\hat{\theta}_{k-1}$ và có độ phân biệt lớn nhất.

Thuật toán tìm kiếm nhị phân.

Cho trước một giá trị a , tìm trong tập hợp b_1, b_2, \dots, b_n giá trị gần nhất với giá trị a . Cách làm đơn giản nhất là ta tính sai số giữa a với tất cả các giá trị trong tập hợp, sau đó tìm sai số nhỏ nhất. Cách làm này rất mất thời gian vì độ phức tạp của thuật toán là $O(n)$. Thuật toán tìm kiếm nhị phân cho phép ta tìm kiếm một cách nhanh hơn. Ta chỉ cần so sánh giá trị a với giá trị trung tâm (là giá trị nằm ở vị trí giữa) trong dãy. Nếu a nhỏ hơn giá trị trung tâm thì ta tìm kiếm trong nửa trái của dãy, nếu a lớn hơn giá trị trung tâm thì ta tiếp tục tìm kiếm trong nửa phải của dãy, nếu a xấp xỉ giá trị trung tâm thì ta lấy luôn giá trị trung tâm. Độ phức tạp của thuật toán này là $O(\log n)$. Ta có giả thuật như sau:

```

Input: a, b1, b2, ..., bn
Output: b (estimate a)
L := 0
R := n
Do while (L < R)
    m := [L + R/2]
    if (a = bm) then
        b = bm
    else
        if (a < bm) then
            R := m - 1
        else
            L := m + 1

```

```
endif
endif
enddo
```

Điều kiện dừng

Nhiều tiêu chí dừng có thể được thiết lập: số lượng câu hỏi tối đa hoặc tối thiểu, giới hạn thời gian làm bài, và độ chính xác của ước tính. Khi một trong các điều kiện dừng được đáp ứng, bài kiểm tra kết thúc; nếu không, một câu hỏi mới sẽ được đưa ra cho các thí sinh.

Sai số chuẩn:

Khi thí sinh trả lời, ngoài khả năng của thí sinh là ước lượng, cần tính sai số chuẩn của ước lượng đó, sai số này cho phép ta đo độ chính xác của ước lượng. Chúng ta sử dụng công thức sai số chuẩn sau đây [3].

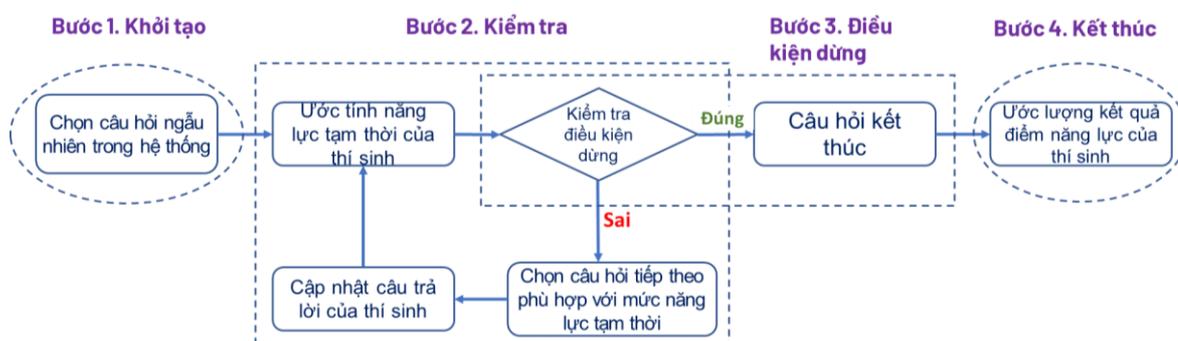
$$ss = \frac{1}{\sqrt{1 + \sum I(\theta^t)}}$$

Một bài trắc nghiệm thích ứng trên hệ thống UEd-CAT bao gồm bốn bước, cụ thể:

1. Bước khởi tạo: Hệ thống chọn ngẫu nhiên 1 câu hỏi trong ngân hàng câu hỏi. Sau khi có thông tin về câu trả lời của thí sinh, năng lực của thí sinh được ước lượng tạm thời ký hiệu là $\hat{\theta}$.

2. Bước kiểm tra: Hệ thống chọn 1 câu hỏi kiểm tra năng lực tạm thời $\hat{\theta}$. Sau đó dựa trên câu trả lời của thí sinh, năng lực $\hat{\theta}$ sẽ được cập nhật ở mức độ cao hơn hoặc thấp hơn.

3. Bước dừng và kết thúc: Tiêu chí dừng sẽ được kiểm tra cho đến khi thỏa mãn, vòng lặp kết thúc và ước tính năng lực cuối cùng của thí sinh (bước kết thúc). Nếu không, một câu hỏi khác sẽ được chọn dựa trên θ . Vòng lặp cập nhật năng lực cứ tiếp tục lặp đi lặp lại.



Hình 2.5: Các bước của một bài trắc nghiệm thích ứng trên hệ thống UEd-CAT (Lê Thái Hưng & Nguyễn Thái Hà, 2020)

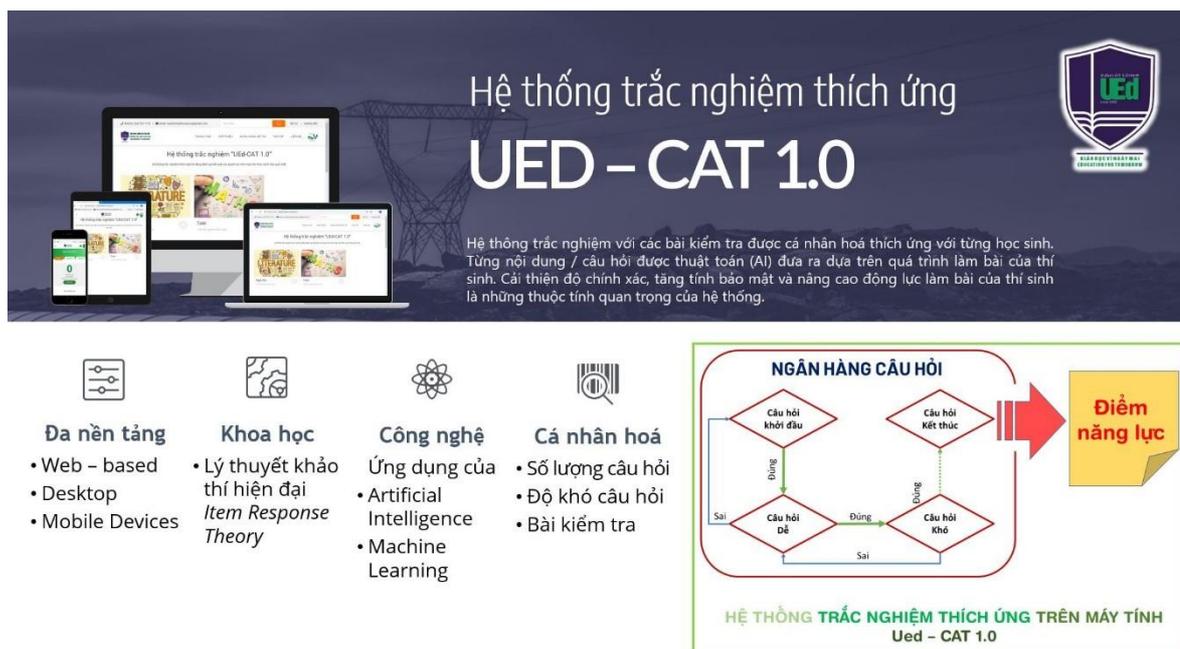
Từ các kết quả nghiên cứu được công bố, có thể thấy rằng hệ thống UEd-CAT đảm bảo các tính năng vượt trội của trắc nghiệm thích ứng trên máy tính với số lượng câu hỏi ít hơn, theo đó thời gian làm bài ngắn hơn nhưng độ chính xác vẫn được đảm bảo, cùng với đó là sự đón nhận của học sinh với trắc nghiệm thích ứng (Le & Nguyen, 2021).

Hiện tại, các đề thi trắc nghiệm thích ứng đánh giá năng lực vật lý, hóa học, sinh học của học sinh các khối lớp đang được thử nghiệm và nghiên cứu. Bên cạnh đó, các tính năng bảo mật, khả năng vận hành thuận tiện trong hệ sinh thái dạy học thích ứng UED – ALS: Adaptive Learning System cũng được tiếp tục đầu tư phát triển. Nhờ đó, hệ thống UEd-CAT hứa hẹn sẽ mang lại một công cụ dạy và học cũng như tự đánh giá kết quả học tập hiệu quả và tạo hứng thú cho cả người dạy và người học trong thời gian tới, đồng thời góp phần hiện

thực hoá xu thế cá nhân hóa quá trình học tập cũng như đổi mới dạy học và kiểm tra đánh giá tại Việt Nam.

2.4.2. Bài trắc nghiệm song ngữ đánh giá từ vựng tiếp nhận tiếng Anh

2.4.2.1. Phiên bản song ngữ của bài kiểm tra từ vựng tiếp nhận tiếng Anh



Hình 2.6: Tính năng của Ued-CAT

(Lê Thái Hưng & Nguyễn Thái Hà, 2020)

Bài trắc nghiệm song ngữ đánh giá từ vựng tiếp nhận tiếng Anh luận án xây dựng dựa trên bài trắc nghiệm tiếng Anh thông dụng NGSST của Stoeckel và Bennett (2015). Đây không phải là nghiên cứu đầu tiên phát triển phiên bản song ngữ của bài kiểm tra đánh giá kiến thức từ vựng tiếp nhận tiếng Anh. Trước đây, VST, một trong những bài kiểm tra đánh giá kiến thức từ vựng tiếp nhận tiếng Anh phổ biến nhất, đã tồn tại với một số phiên bản song ngữ, bao gồm tiếng Ả Rập, tiếng Gujarati, tiếng Nhật, tiếng Hàn, tiếng Quan Thoại, tiếng

Nga, tiếng Tamil, tiếng Thái và tiếng Việt, tất cả đều có tại <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests>.

uniform: He has his **uniform**.

- a. clothes for everyone in a company to wear
- b. suggestion for a business plan
- c. thing that makes you less successful
- d. reason to do something

Hình 2.7: Câu hỏi ví dụ trong NGSLT

Trong các phiên bản song ngữ, các từ được kiểm tra là tiếng Anh và các lựa chọn đáp án được xây dựng bằng ngôn ngữ mẹ đẻ của người làm bài kiểm tra. Trên đây là một câu hỏi ví dụ từ NGSLT đơn ngữ để minh họa lợi ích của phiên bản song ngữ. Qua ví dụ này có thể thấy, để chọn nghĩa đúng cho từ được kiểm tra, thí sinh cần hiểu tất cả các phương án cho sẵn, trong trường hợp này, liên quan đến kỹ năng đọc các phương án dài và kiến thức ngữ pháp về mệnh đề quan hệ và cụm danh từ phức tạp với giới từ và động từ nguyên mẫu. Một gợi ý đã được đưa ra và được chứng minh là vừa làm tăng giá trị của bài kiểm tra đánh giá kiến thức từ vựng tiếp nhận vừa vượt qua hạn chế tiềm năng của những yếu tố gây nhiễu không phải từ từ vựng, đó là xây dựng các phương án lựa chọn ngắn bằng ngôn ngữ mẹ đẻ của thí sinh (Stewart, 2009; Nguyen & Nation, 2011; Karami, 2012; Elgort, 2013).

Mặc dù NGSLT mới được phát triển và xác trị nhưng một phiên bản tiếng Nhật song ngữ đã được tạo ra bởi Stoeckel và cộng sự (2018). Nghiên cứu đã cung cấp bằng chứng đáng tin cậy và khẳng định sự phù hợp cũng như lợi ích của phiên bản song ngữ, chủ yếu khi NGSLT nhắm đến những người làm bài kiểm tra có kiến thức ngữ pháp và kỹ năng đọc hạn chế.

justify: We cannot **justify** this.

- a. hit this hard
- b. follow this
- c. exchange this for something else
- d. show that this is right

justify: We cannot **justify** this.

- a. 打ち砕く
- b. 追求する
- c. 変更する
- d. 正当化する

Hình 2.8: Câu hỏi ví dụ của NGSLT
trong bài thi song ngữ tiếng Anh và tiếng Nhật

Cùng với phiên bản song ngữ Anh – Nhật được công bố bởi Stoeckel và cộng sự (2018), nghiên cứu hiện tại tiếp tục hướng đề xuất bởi nhóm học giả thiết kế NGSLT bằng cách phát triển và đánh giá hiệu quả một bài kiểm tra song ngữ tiếng Anh và tiếng Việt dựa trên các bước xây dựng được báo cáo trong các nghiên cứu liên quan trước đây.

Bài thi NGSLT bao gồm 100 câu hỏi trắc nghiệm. Định dạng câu hỏi giống như định dạng của VST, được cho là ít thách thức hơn so với định dạng ghép nối của VLT (Kamimoto, 2014; Culligan, 2015). Trong mỗi câu hỏi, một từ được đưa ra cùng với một câu sử dụng từ đó trong một ngữ cảnh cụ thể. Mỗi câu hỏi có bốn lựa chọn trả lời, trong đó có ba lựa chọn nhiễu và một đáp án đúng. Kết quả kiểm tra của NGSLT cung cấp thông tin về các cấp độ từ vựng giống như VLT, được cho là đưa ra nhiều gợi ý có giá trị hơn cho thí sinh trong việc phát triển vốn từ vựng so với VST (Schmitt và cộng sự, 2020). Trong bối cảnh dạy và học tiếng Anh như một ngoại ngữ ở Việt Nam, bài thi NGSLT có thể coi là một lựa chọn phù hợp cho cả giáo viên và học sinh, giúp họ xác định mục tiêu từ vựng trong chương trình dạy và học cũng như lựa chọn tài liệu dạy và học phù hợp (Koizumi, 2015; Stoeckel và cộng sự, 2018).

2.4.2.2. Bảng đặc tả bài kiểm tra từ vựng song ngữ

Xây dựng bảng đặc tả gồm các thông số đặc tính của bài kiểm tra được coi là bước đầu tiên trong quá trình xây dựng đề kiểm tra vì chúng là “bản thiết kế chi tiết” dùng trong cả quá trình (Davidson & Lynch, 2002, tr. 1).

Bảng đặc tả bài kiểm tra bao gồm thông tin cần thiết của bài kiểm tra, bao gồm mục đích của bài kiểm tra, đối tượng hướng tới, nội dung, cấu trúc và định dạng bài kiểm tra. Bảng đặc tả của bài kiểm tra song ngữ từ vựng tiếng Anh thông dụng trong nghiên cứu được trình bày trong Bảng 2.6. Bảng đặc tả bài kiểm tra từ vựng song ngữ kế thừa các đặc tính của bài kiểm tra NGSLT, bên cạnh những bổ sung về việc viết câu hỏi trong phiên bản song ngữ theo các hướng dẫn Nguyen và Nation (2011) và Stoeckel và đồng sự (2018) chỉ ra trong nghiên cứu của mình.

Bảng 2.6: Bảng đặc tả bài kiểm tra song ngữ đánh giá từ vựng tiếp nhận tiếng Anh

Mục đích	Đánh giá từ vựng tiếp nhận tiếng Anh					
Đối tượng kiểm tra	Người học ngoại ngữ tiếng Anh ở Việt Nam, trình độ tiếng Anh bậc 1-2-3, tiếng Việt là ngôn ngữ mẹ đẻ					
Danh sách từ kiểm tra	100 từ được lấy từ Danh sách từ vựng tiếng Anh thông dụng mới (Browne, 2013), danh sách gồm 5 mức tần suất, mỗi mức có 560 từ. Tỉ lệ câu hỏi chia theo bậc thông dụng của danh sách từ:					
	Mức tần suất	1	2	3	4	5
	Số câu	20	20	20	20	20
	Tỉ lệ	20%	20%	20%	20%	20%
Chỉ dẫn làm bài	- bằng tiếng Việt, yêu cầu thí sinh chọn phương án gần nghĩa nhất với từ được kiểm tra. - Có một ví dụ minh họa					

Thời gian làm bài	Không giới hạn
Định dạng câu hỏi	<p>Câu hỏi nhiều lựa chọn</p> <ul style="list-style-type: none"> - Một từ tiếng Anh được đưa ra, sau đó là một câu sử dụng từ đó trong ngữ cảnh cụ thể. - Có 4 lựa chọn, trong đó có 1 lựa chọn đúng và 3 lựa chọn gây nhiễu bằng tiếng Việt
Kỹ thuật dịch/viết câu hỏi song ngữ	<ul style="list-style-type: none"> - Với câu đề dẫn sử dụng từ kiểm tra: câu không có ngữ cảnh cụ thể, chỉ có tác dụng giới hạn nghĩa của từ được kiểm tra trong nhiều ngữ của từ, không chứa gợi ý để thí sinh đoán được đáp án. - Với các lựa chọn tiếng Việt <ul style="list-style-type: none"> • Diễn đạt nghĩa tự nhiên, thể hiện được chính xác nghĩa của từ được kiểm tra • Nhất quán về loại từ, sắc thái nghĩa (tích cực hay tiêu cực) trong các lựa chọn • Cân đối số lượng từ sử dụng trong các lựa chọn, không có lựa chọn khác biệt rõ rệt về số lượng từ với các lựa chọn còn lại. • Không sử dụng các từ vay mượn có thể chứa gợi ý để thí sinh đoán nghĩa. • Trong trường hợp dịch từ bài kiểm tra NGSLT: lựa chọn đúng phải dịch trực tiếp từ kiểm tra trong tiếng Việt, không dịch định nghĩa tiếng Anh trong lựa chọn đúng của câu hỏi gốc, ví dụ: từ được kiểm tra.
Tính điểm	1 điểm cho mỗi câu trả lời đúng
Thông báo kết quả	Tổng số điểm trên thang 100

2.4.2.3. Biên soạn và thử nghiệm đề song ngữ

Tuân theo đặc tả bài kiểm tra, đặc biệt là kỹ thuật viết câu hỏi trong phiên bản song ngữ, 100 câu hỏi đã được chuyển dịch từ bài kiểm tra NGSLT để xây dựng phiên bản song ngữ của NGSLT (Phụ lục 4).

Sau đây là một câu hỏi ví dụ từ bài kiểm tra song ngữ:

<i>Câu hỏi trong bài kiểm tra NGSLT</i>	<i>Câu hỏi trong bài kiểm tra song ngữ</i>
<p>uniform: He has his uniform.</p> <p>a. clothes for everyone in a company to wear</p> <p>b. suggestion for a business plan</p> <p>c. thing that makes you less successful</p> <p>d. reason to do something</p>	<p>uniform: He has his uniform.</p> <p>a. đồng phục</p> <p>b. đề xuất kinh doanh</p> <p>c. trở ngại</p> <p>d. lý do</p>

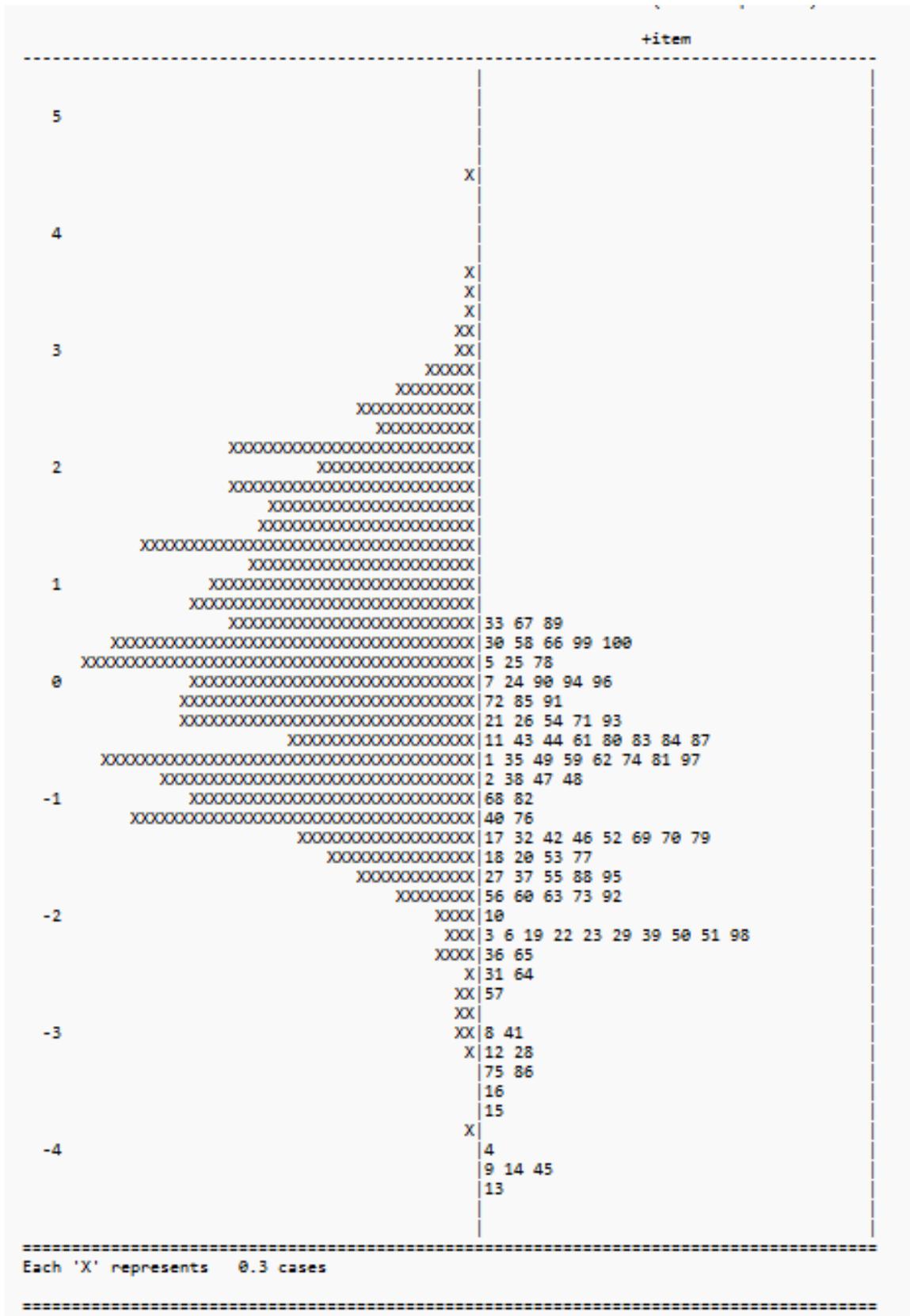
Hình 2.9: Câu hỏi ví dụ trong NGSLT song ngữ tiếng Anh và tiếng Việt

Đánh giá độ tin cậy của đề mẫu

Sau quá trình thử nghiệm với 225 thí sinh, dữ liệu thu được được phân tích với phần mềm Conquest. Hệ số Alpha (Coefficient Alpha) và hệ số tin cậy độc lập (Separation Reliability) của đề mẫu đều rất cao. Các hệ số này cao cho thấy các đề thiết kế có độ tin cậy cao và các tham số câu hỏi có tính độc lập tốt. Hệ số cao như vậy có thể giải thích bằng việc bài kiểm tra đánh giá đơn năng lực – nhận diện nghĩa từ đã cho với dữ kiện là một câu hoàn chỉnh có sử dụng từ nhưng hạn chế khả năng đoán nghĩa từ ngữ cảnh, và mức độ định dạng thống nhất chặt chẽ của các câu hỏi trong đề kiểm tra. Mỗi câu hỏi chỉ kiểm tra một từ duy nhất trong danh sách NGS� và các lựa chọn cũng như đáp án không có sự trùng lặp hay gợi ý cho các câu hỏi khác.

Bảng 2.7: Hệ số Alpha và hệ số tin cậy độc lập của đề mẫu

Hệ số tin cậy độc lập	Hệ số Alpha
0.917	0.9



Hình 2.10: Bản đồ phân bố năng lực và độ khó của đề mẫu

Hình 2.10 là bản đồ phân bố năng lực và độ khó của đề mẫu. Trong bản đồ này, ước tính khả năng của thí sinh được trình bày ở bên trái và mức độ khó của câu hỏi được trình bày ở bên phải. Từ bản đồ này, mức độ năng lực có thể đo trong khoảng -4 đến 4, điều này cho thấy rằng bài kiểm tra có thể đo được khoảng năng lực rộng của thí sinh.

Loại các câu hỏi không phù hợp với mô hình

Về mức độ phù hợp của câu hỏi với mô hình, chỉ số Weighted Fit được sử dụng để phát hiện các câu hỏi không phù hợp. Bảng 2.8 minh họa kết quả phân tích của sáu câu bị loại của đề mẫu do có MNSQ không nằm trong phạm vi CI và thống kê T tương ứng của chúng có giá trị tuyệt đối vượt quá 2,0.

Bảng 2.8: Các câu hỏi không phù hợp với mô hình của đề mẫu

VARIABLES		UNWEIGHTED FIT			WEIGHTED FIT			
-----		-----	-----	-----				
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
5 5	-1.134	0.178	1.32 (0.81, 1.19)	3.1	1.25 (0.82, 1.18)	2.5		
7 7	-1.056	0.177	1.61 (0.81, 1.19)	5.4	1.29 (0.82, 1.18)	3.0		
21 21	-2.040	0.211	0.44 (0.81, 1.19)	-7.2	0.73 (0.76, 1.24)	-2.4		
25 25	-0.768	0.176	1.21 (0.81, 1.19)	2.1	1.22 (0.83, 1.17)	2.4		
33 33	-0.160	0.172	1.71 (0.80, 1.20)	5.9	1.48 (0.84, 1.16)	5.3		
99 99	-1.439	0.212	0.57 (0.79, 1.21)	-4.8	0.75 (0.77, 1.23)	-2.3		

Phân loại và chỉnh sửa câu hỏi

Để tăng chất lượng câu hỏi, nhóm tác giả sử dụng kết quả phân tích của Conquest với từng câu hỏi. Tất cả 94 câu hỏi phù hợp với mô hình được phân chia thành hai nhóm: nhóm câu hỏi tốt và nhóm câu hỏi cần chỉnh sửa.

Nhóm câu hỏi tốt phù hợp với mô hình Rasch, có độ khó và độ phân biệt chấp nhận được theo lý thuyết khảo thí cổ điển cùng với các phương án gây

nhiều đạt yêu cầu. Bảng 2.9 là kết quả phân tích của một câu hỏi tốt của đề mẫu – câu hỏi 66.

Với lý thuyết khảo thí cổ điển, độ khó của câu hỏi được tính là 0,69, nằm trong khoảng 0,25-0,75; 69,46% thí sinh trả lời đúng câu hỏi này. Độ phân biệt rất tốt, $d = 0,5$, cho thấy câu hỏi có thể phân biệt giữa nhóm thí sinh có trình độ cao và nhóm thí sinh có năng lực thấp. Hệ số tương quan (PT BIS) cho chúng ta thấy những phương án gây nhiễu có chỉ số âm và phương án đúng có chỉ số dương. Các phương án đưa ra ở câu 66 có giá trị đánh giá năng lực của thí sinh.

Bảng 2. 9: Kết quả phân tích của câu hỏi 66

Item 66							

item:66 (66)							
Cases for this item	177	Discrimination	0.50				
Item Threshold(s):	-1.06	Weighted MNSQ	1.10				
Item Delta(s):	-1.06						

Label	Score	Count	% of tot	Pt Bis	t (p)	WLEAvg:1	WLE SD:1

a	0.00	32	18.08	-0.27	-3.65(.000)	-0.93	1.44
b	0.00	3	1.69	-0.30	-4.11(.000)	-2.73	1.54
c	0.00	19	10.73	-0.28	-3.92(.000)	-1.24	1.27
d	1.00	123	69.49	0.50	7.57(.000)	0.60	1.34
=====							

Kết quả này được khẳng định thông qua đường cong đặc trưng của câu hỏi. Nhìn Hình 2.11 ta có thể thấy đường cong đặc trưng của câu hỏi (đường nét đứt) bám sát đường cong kỳ vọng của câu hỏi, xuất phát từ điểm xác suất gần gốc tọa độ đi lên, cho thấy câu hỏi thiết kế phù hợp để kiểm tra năng lực của thí sinh.

Nhóm câu hỏi cần chỉnh sửa gồm các câu hỏi phù hợp với mô hình Rasch, có độ khó và độ phân biệt chấp nhận được theo lý thuyết khảo thí cổ điển. Sau khi phân tích đề mẫu song ngữ, nhóm câu hỏi này được chỉnh sửa phương án nhiễu để tham gia vào các bước nghiên cứu tiếp theo.



Hình 2.11: Đường cong đặc trưng của câu hỏi 66

2.4.3. Bảng câu hỏi xin ý kiến chuyên gia

Bảng câu hỏi xin ý kiến chuyên gia có thể tìm thấy ở Phụ lục 1, bao gồm 3 phần. Phần 1 là thông tin của chuyên gia, bao gồm giới tính, chứng chỉ bằng cấp liên quan đến trình độ tiếng Anh, giảng dạy tiếng Anh và công tác ra đề kiểm tra. Phần 2 bao gồm các tiêu chí đánh giá bộ câu hỏi thô, bao gồm mục đích, đối tượng, nội dung, cấu trúc của đề kiểm tra, gợi ý chỉnh sửa. Phần cuối là phần câu hỏi mở để các chuyên gia đưa ra ý kiến về các vấn đề cần chỉnh sửa cũng như đưa ra các gợi ý để hoàn thiện bộ câu hỏi thô.

2.4.4. Bảng câu hỏi khảo sát

Sau khi thí sinh được trực tiếp trải nghiệm đề trắc nghiệm thích ứng trên hệ thống, thí sinh được mời tham gia trả lời khảo sát để phản ánh về trải nghiệm của mình. Bảng câu hỏi khảo sát được thiết kế với ba phần. Phần thứ nhất là

thông tin cá nhân của người tham gia bao gồm giới tính, tự đánh giá lượng từ, trải nghiệm với trắc nghiệm thích ứng. Phần thứ hai yêu cầu người tham gia khảo sát lựa chọn mức độ đồng ý với các nhận định về bài kiểm tra thích ứng từ vựng mà họ đã thực hiện, bao gồm thời gian, thao tác thực hiện, tính chính xác, sự phù hợp với năng lực thí sinh, căng thẳng trong quá trình làm bài, khả năng hạn chế gian lận, thông báo kết quả tức thời, gợi ý học tập, công cụ tự đánh giá và công cụ hỗ trợ học tập. Phần cuối là phần mở để người tham gia có thể chia sẻ những nhận định tổng quát nhất về trải nghiệm của mình, những gì hài lòng và không hài lòng nhất cũng như mong muốn thực hiện các bài kiểm tra thích ứng khác trong tương lai. Những nhận định tích cực và tiêu cực của người tham gia khảo sát đều được tôn trọng nhằm mang lại cái nhìn khách quan về bài kiểm tra thích ứng, từ đó góp phần phát triển những nhận định và đề xuất phù hợp của cả luận án.

Do tổng thể khảo sát là các thí sinh có trải nghiệm trực tiếp với bài kiểm tra thích ứng trong nghiên cứu, cụ thể là 98 sinh viên, nên nhà nghiên cứu chỉ tiến hành thử nghiệm bảng câu hỏi khảo sát với ba sinh viên để đảm bảo chất lượng ngôn ngữ và khả năng khai thác thông tin của bảng câu hỏi khảo sát trước khi tiến hành điều tra khảo sát chính thức.

2.4.5. Bộ câu hỏi phỏng vấn

Bộ câu hỏi phỏng vấn dựa trên nội dung bộ câu hỏi khảo sát, với nhiều các câu hỏi mở yêu cầu người tham gia phỏng vấn có thể trình bày chi tiết quan điểm, lý giải cũng như nhận định của mình – điều mà điều tra khảo sát chưa thực hiện được.

Cấu trúc của bộ câu hỏi phỏng vấn gồm bốn phần: (1) thông tin cá nhân, (2) những trải nghiệm trước đây với trắc nghiệm thích ứng, (3) ý kiến về bài kiểm tra thích ứng từ vựng đã thực hiện, bao gồm thời gian, thao tác thực hiện, tính chính xác, sự phù hợp với năng lực thí sinh, căng thẳng trong quá trình làm

bài, khả năng hạn chế gian lận, thông báo kết quả tức thời, gợi ý học tập, công cụ tự đánh giá và công cụ hỗ trợ học tập, và (4) nhận định chung về bài kiểm tra. Các câu hỏi mở hỏi thêm thông tin, yêu cầu chi tiết lý giải được thêm vào để làm rõ hơn về trải nghiệm của người tham gia với việc học ngoại ngữ tiếng Anh cũng như trải nghiệm thích ứng nói chung và đến trải nghiệm thích ứng đánh giá từ vựng tiếng Anh thông dụng. Các ý kiến về trải nghiệm cũng kèm thêm các câu hỏi mở để người tham gia có thể lý giải cụ thể về ý kiến của mình từ đó đưa ra những đề xuất hỗ trợ việc nâng cao chất lượng đề và hệ thống trắc nghiệm thích ứng cũng như đề xuất về việc sử dụng trắc nghiệm thích ứng hiệu quả nâng cao chất lượng giáo dục nói chung.

2.5. Các vấn đề về đạo đức nghiên cứu

Vấn đề đạo đức nghiên cứu được cân nhắc ở tất cả các giai đoạn nghiên cứu, từ lên kế hoạch lấy mẫu, gửi lời mời tham gia, tiến hành thử nghiệm để thu thập số liệu. Tất cả đối tượng tham gia nghiên cứu đều được thông báo về mục đích nghiên cứu, hướng dẫn cách thức tham gia, xác nhận tự nguyện tham gia đồng thời quyền được rút lui không tham gia nghiên cứu.

Cam kết của người tham gia

- Em xác nhận rằng em đã được cung cấp và hiểu các thông tin về nghiên cứu đang thực hiện. *
- Em khẳng định rằng em đồng ý tham gia vào nghiên cứu và việc tham gia của em là tự nguyện.
- Em đồng ý rằng dữ liệu thu thập được trong quá trình nghiên cứu chỉ được sử dụng với mục đích nghiên cứu và đánh giá nghiên cứu.

Xác nhận cam kết

Hình 2.12: Xác nhận của người tham gia nghiên cứu

Các thông tin cá nhân và câu trả lời được cam kết bảo mật và chỉ sử dụng cho mục đích nghiên cứu. Việc báo cáo kết quả được tiến hành chính xác và trung thực, nhà nghiên cứu thay đổi danh tính của tất cả các đối tượng tham gia nghiên cứu trong quá trình báo cáo, đảm bảo không ảnh hưởng đến danh dự và quyền lợi của bất cứ cá nhân nào.

2.6. Kết chương

Chương 2 đã trình bày phương pháp nghiên cứu của luận án với các nội dung chính: thiết kế nghiên cứu, phương pháp nghiên cứu, quá trình lấy mẫu, công cụ nghiên cứu, phân tích dữ liệu và các vấn đề về đạo đức nghiên cứu. Từ các nội dung này, chuyên đề hoàn thành việc xây dựng bảng đặc tả đề trắc nghiệm đánh giá kiến thức từ vựng tiếp nhận tiếng Anh và phân tích đề trắc nghiệm mẫu. Bảng đặc tả và việc phân tích đề mẫu sẽ được sử dụng trong giai đoạn tiếp theo để chuẩn hóa ngân hàng câu hỏi của hệ thống trắc nghiệm thích ứng và so sánh kết quả thực nghiệm của đề truyền thống và đề trắc nghiệm thích ứng. Các công cụ nghiên cứu bao gồm bảng câu hỏi xin ý kiến chuyên gia, bảng câu hỏi khảo sát thí sinh, bộ câu hỏi phỏng vấn thí sinh đã được thiết kế và hoàn thiện, sẵn sàng cho giai đoạn thu thập và phân tích dữ liệu.

CHƯƠNG 3: KẾT QUẢ NGHIÊN CỨU VÀ BÀN LUẬN

Kết quả nghiên cứu của luận án được chia thành hai nội dung chính, tương ứng với hai câu hỏi nghiên cứu luận án đã đề ra: chuẩn hóa ngân hàng câu hỏi trắc nghiệm thích ứng và đánh giá hiệu quả của bài kiểm tra thích ứng đánh giá từ vựng tiếp nhận tiếng Anh dành cho người học tiếng Anh ở Việt Nam.

3.1. Chuẩn hóa ngân hàng câu hỏi trắc nghiệm thích ứng từ vựng tiếp nhận tiếng Anh

3.1.1. Biên soạn và chỉnh sửa bộ câu hỏi thô

Quá trình thiết kế bộ câu hỏi thô được chia làm hai bước chính. Bước một là biên soạn bộ câu hỏi thô từ bảng đặc tả bài trắc nghiệm song ngữ đánh giá từ vựng tiếp nhận tiếng Anh đã được trình bày trong chương 2, bao gồm mục đích kiểm tra, đối tượng kiểm tra, danh sách từ nguồn, định dạng câu hỏi và hướng dẫn viết câu hỏi. Bước thứ hai là xin ý kiến chuyên gia để chỉnh sửa bộ câu hỏi thô.

Theo Thompson và Weiss (2011), một ngân hàng có từ 400 đến 500 câu hỏi có khả năng đảm bảo chất lượng cao, trong đó có 100 câu hỏi sẽ được luân chuyển ở các vị trí khác nhau khi thí sinh làm bài kiểm tra trắc nghiệm thích ứng. Một số các nghiên cứu phát triển đề kiểm tra trắc nghiệm thích ứng báo cáo số lượng câu hỏi của ngân hàng câu hỏi thấp hơn, như Kustiyahningsih và Cahyani (2013) và Tseng (2016) có ngân hàng với 180 câu hỏi. Tuy nhiên với hệ thống trắc nghiệm thích ứng của trường ĐHGD, nhóm nghiên cứu đã và đang hướng tới chuẩn hóa ngân hàng 500 câu hỏi cho các nội dung kiểm tra như ngữ văn và toán. Tuân thủ yêu cầu của nhóm nghiên cứu, nghiên cứu này cũng tiến hành với mục tiêu số lượng câu hỏi của ngân hàng là 500.

Với 100 câu hỏi chuyển dịch từ đề tiếng Anh gốc, nhà nghiên cứu tiếp tục tiến hành lựa chọn thêm 452 từ từ danh sách từ vựng tiếng Anh thông dụng mới của Browne (2013) (Phụ lục 6) và viết câu hỏi theo bảng đặc tả với lưu ý đảm bảo các câu hỏi thô có độ phủ cả năm mức độ tần suất của danh sách từ, mỗi mức độ có số câu hỏi lớn hơn 100. Bảng 3.1 tổng hợp 552 câu hỏi phân theo mức độ tần suất:

Bảng 3.1: Số câu hỏi thô theo mức độ tần suất

Mức độ tần suất	1	2	3	4	5
Thứ hạng trong danh sách từ	1 - 560	561- 1120	1121- 1680	1681 - 2240	2241 - 2801
Số câu hỏi	110	111	110	110	111

Trong bước tiếp theo, bộ câu hỏi thô được chia thành bảy đề tuân thủ bảng đặc tả của bài kiểm tra song ngữ từ vựng tiếp nhận tiếng Anh, có thiết kế câu hỏi neo để thực hiện bước cân bằng đề sau này với mục đích chuẩn hóa ngân hàng câu hỏi.

Bảng 3.2: Thông tin thẩm định đề của nhóm chuyên gia

Đề	Chuyên gia 1	Chuyên gia 2	Chuyên gia 3	Chuyên gia 4	Chuyên gia 5	Chuyên gia 6	Chuyên gia 7
1	X		X		X		
2		X		X		X	
3	X		X			X	
4		X		X	X		
5	X	X					X
6			X		X		X
7				X		X	X

Các đề này được chuyển đến cho các chuyên gia tiến hành thẩm định, cùng với bảng đặc tả cũng như các thông tin của nghiên cứu. Mỗi đề sẽ được

thẩm định bởi 3 chuyên gia để đảm bảo tính khách quan. Thông tin thẩm định đề của các chuyên gia được tóm lược trong Bảng 3.2.

Kết quả đánh giá của các chuyên gia tham gia nghiên cứu được tổng hợp theo hai nội dung chính, đánh giá chung về các đề và đề xuất chỉnh sửa câu hỏi. Thứ nhất, với các tiêu chí đánh giá đề kiểm tra bao gồm mục đích kiểm tra, nội dung kiểm tra, cấu trúc đề kiểm tra và đối tượng kiểm tra, các đề đều được nhóm chuyên gia đánh giá ở mức tốt và rất tốt, điểm trung bình các tiêu chí ở khoảng 4,42 đến 4,75, cho thấy chất lượng các đề được các chuyên gia đánh giá cao, đáng tin cậy (Bảng 3.3). Thứ hai, các nhóm chuyên gia cũng tiến hành đánh giá từng câu hỏi với kiến thức và kinh nghiệm chuyên môn của mình. Các đề xuất được đưa ra bao gồm việc chỉnh sửa câu sử dụng từ trong các câu hỏi và các phương án nhiễu. Bảng 3.4 là ví dụ một câu hỏi được nhóm chuyên gia đề xuất chỉnh sửa. Sau khi nhận các đề xuất, nhà nghiên cứu tiến hành chỉnh sửa theo các đề xuất trước khi tiến hành thử nghiệm với đối tượng thí sinh.

Bảng 3.3: Tổng hợp đánh giá của nhóm chuyên gia

Đề	Mục đích kiểm tra rõ ràng	Nội dung kiểm tra chính xác	Cấu trúc đề kiểm tra logic	Đối tượng kiểm tra phù hợp	Trung bình
1	4.67	4.67	5.00	4.00	4.67
2	4.33	5.00	4.33	4.00	4.50
3	5.00	4.33	5.00	5.00	4.67
4	4.33	4.33	4.67	4.00	4.42
5	4.67	4.67	5.00	4.00	4.75
6	4.67	4.00	4.67	4.00	4.42
7	4.67	4.67	4.67	5.00	4.58

Bảng 3.4: Ví dụ câu hỏi chỉnh sửa theo ý kiến chuyên gia

Câu hỏi	Nhận xét của chuyên gia	Chỉnh sửa
<p>potential: It is one of potential risks.</p> <p>a. có thể b. lớn c. chính d. có thật</p>	<p>Chỉnh lại phương án a (gợi ý: tiềm tàng)</p>	<p>potential: It is one of potential risks.</p> <p>a. tiềm tàng b. lớn c. chính d. có thật</p>

3.1.2. Đánh giá độ tin cậy của các đề thử nghiệm

Các đề được thiết kế từ bộ câu hỏi thô được tiến hành thử nghiệm đồng thời với sự tham gia của 1619 sinh viên. Tuy nhiên, sau quá trình làm sạch dữ liệu, nghiên cứu đưa đến quyết định chỉ lựa chọn kết quả bài làm của 1081 sinh viên, đảm bảo mẫu lớn hơn 150 với từng đề và kết quả bài làm của mẫu được chọn có đủ dữ liệu của cả 100 câu hỏi trong đề thử nghiệm. Số lượng các bài làm bị loại bỏ do chưa hoàn thành việc trả lời đầy đủ 100 câu trong bài kiểm tra thử nghiệm. Số lượng bị loại bỏ tuy khá lớn nhưng là cần thiết để đảm bảo tính chính xác trong dữ liệu của từng câu hỏi trong bài kiểm tra.

Bảng 3.5: Cơ mẫu được chọn phân tích của bảy bài kiểm tra thử nghiệm

Đề	1	2	3	4	5	6	7	Tổng
Cơ mẫu được chọn phân tích	153	151	155	150	151	168	153	1081

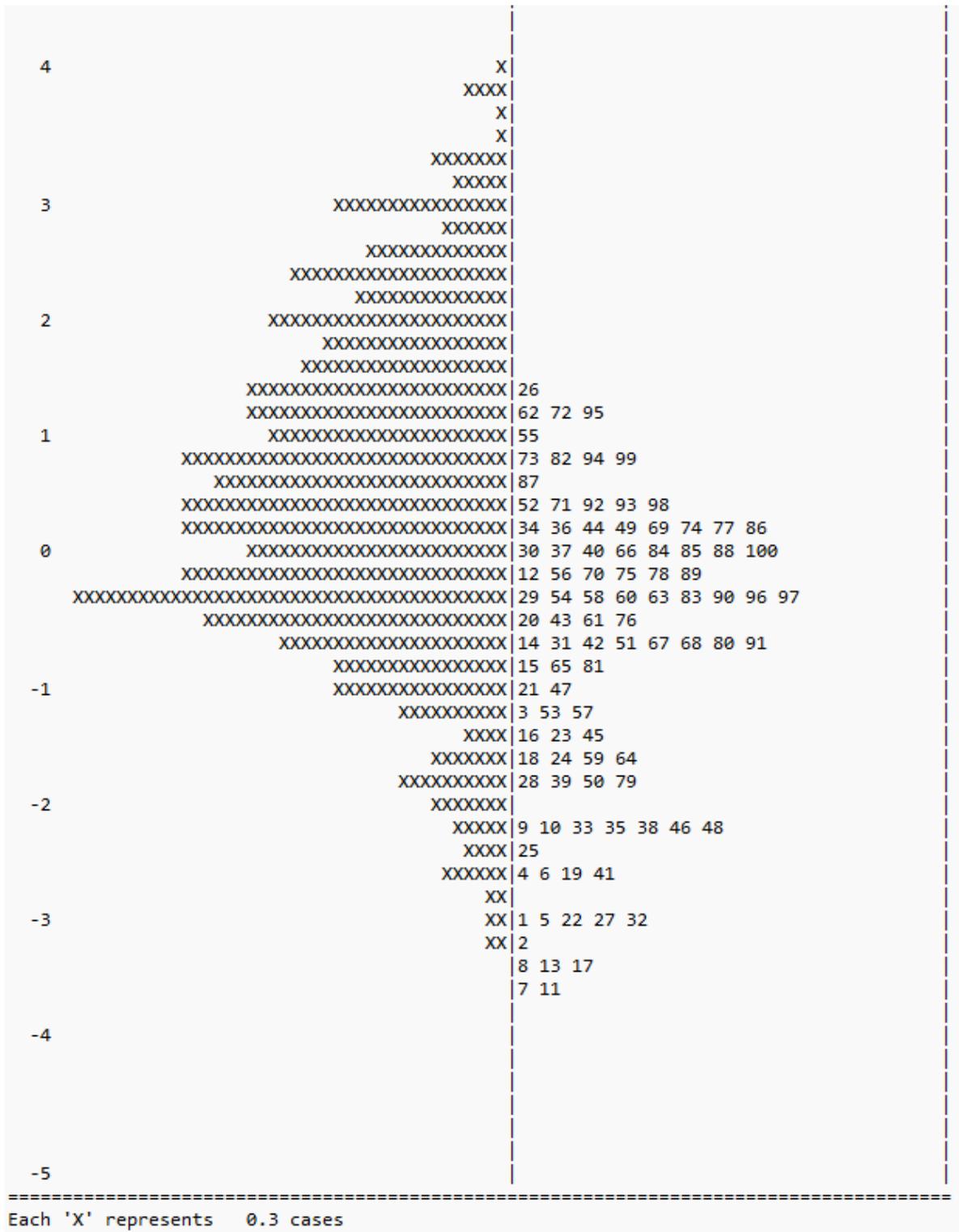
Conquest đã được sử dụng để thực hiện phân tích các đề kiểm tra. Hệ số Alpha và hệ số tin cậy độc lập (Separation Reliability) của cả bảy đề đều trên 0.9. Các hệ số này cao cho thấy các đề thiết kế có độ tin cậy cao và các tham số câu hỏi có tính độc lập tốt. Kết quả này đồng nhất với đánh giá của nhóm

chuyên gia, cho thấy chất lượng của các bài trắc nghiệm được thử nghiệm, cũng như chất lượng của bộ câu hỏi thô nói chung.

Bảng 3.6: Độ tin cậy theo đề

Đề	Hệ số tin cậy độc lập	Hệ số Alpha
1	0.978	0.92
2	0.968	0.93
3	0.958	0.95
4	0.964	0.95
5	0.972	0.95
6	0.960	0.96
7	0.963	0.97

Hình 3.1 là bản đồ phân bố năng lực và độ khó của Đề 7, một trong số bảy đề được chọn để minh họa (kết quả phân tích của các đề còn lại có thể xem trong Phụ lục 4). Trong bản đồ này, ước tính khả năng của thí sinh được trình bày ở bên trái và mức độ khó của câu hỏi được trình bày ở bên phải. Từ bản đồ này, mức độ năng lực có thể đo trong khoảng -4 đến 4, điều này cho thấy rằng bài kiểm tra có thể đo được nhiều loại khả năng của người dự thi. Độ khó của câu hỏi cần được chú ý nhiều hơn khi lượng câu hỏi có mức độ khó trên 2 còn hạn chế. Cũng đáng lưu ý là một số câu hỏi đo lường mức năng lực dưới -3; tuy nhiên, với mục đích đánh giá kiến thức từ vựng của thí sinh, các mục này vẫn có giá trị biểu thị lượng từ vựng mà thí sinh thu được và đánh giá năng lực thí sinh ở trình độ thấp nhất – thí sinh mới bắt đầu học hoặc ở trình độ sơ cấp.



3.1.3. Loại các câu hỏi không phù hợp với mô hình

Về mức độ phù hợp của câu hỏi với mô hình, chỉ số Weighted Fit được sử dụng để phát hiện các câu hỏi không phù hợp.

```

=====
ConQuest: Generalised Item Response Modelling Software  Fri Feb 03 20:31 2023
TABLES OF RESPONSE MODEL PARAMETER ESTIMATES
=====
TERM 1: item
-----
VARIABLES                UNWEIGHTED FIT        WEIGHTED FIT
-----
item  ESTIMATE ERROR^  MNSQ  CI  T  MNSQ  CI  T
-----
1  1      -4.019  0.513  1.08 ( 0.78, 1.22)  0.7  1.04 ( 0.07, 1.93)  0.2
2  2      -4.321  0.589  0.28 ( 0.78, 1.22) -9.1  0.93 ( 0.00, 2.09)  0.1
3  3      -4.321  0.589  1.71 ( 0.78, 1.22)  5.2  1.01 ( 0.00, 2.09)  0.2
4  4      -4.740  0.717  0.25 ( 0.78, 1.22) -9.7  0.95 ( 0.00, 2.36)  0.2
5  5      -1.408  0.203  0.85 ( 0.78, 1.22) -1.3  0.97 ( 0.80, 1.20) -0.2
...
93 93       0.741  0.185  1.00 ( 0.78, 1.22)  0.0  0.96 ( 0.85, 1.15) -0.5
94 94       0.285  0.177  0.88 ( 0.78, 1.22) -1.1  0.90 ( 0.87, 1.13) -1.6
95 95       0.380  0.178  1.29 ( 0.78, 1.22)  2.4  1.19 ( 0.87, 1.13)  2.7
96 96       1.602  0.217  1.09 ( 0.78, 1.22)  0.8  0.99 ( 0.76, 1.24) -0.1
97 97      -0.117  0.176  0.97 ( 0.78, 1.22) -0.2  0.99 ( 0.88, 1.12) -0.2
98 98      -0.460  0.178  1.07 ( 0.78, 1.22)  0.6  1.02 ( 0.87, 1.13)  0.4
99 99      -1.026  0.189  0.94 ( 0.78, 1.22) -0.5  0.98 ( 0.84, 1.16) -0.2
100 100      1.024  0.192  1.28 ( 0.78, 1.22)  2.3  1.09 ( 0.83, 1.17)  1.0
-----
An asterisk next to a parameter estimate indicates that it is constrained
Separation Reliability = 0.978
Chi-square test of parameter equality = 3685.17, df = 100, Sig Level = 0.000
^ Quick standard errors have been used
=====

```

Hình 3.2: Kết quả phân tích Conquest của Đề 1

Bảng 3.7 minh họa kết quả phân tích của năm câu bị loại của Đề 1 do có MNSQ (sai số bình phương trung bình) không nằm trong phạm vi CI (khoảng tin cậy) và thống kê T tương ứng của chúng có giá trị tuyệt đối vượt quá 2,0.

Sau khi phân tích bảy đề bằng Conquest, có tổng số 30 câu hỏi bị loại do không phù hợp với mô hình phân tích.

VARIABLES		UNWEIGHTED FIT			WEIGHTED FIT			
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
39 39	2.020	0.243	2.86 (0.78, 1.22)		11.0	1.36 (0.70, 1.30)		2.1
54 54	-0.365	0.177	1.23 (0.78, 1.22)		1.9	1.18 (0.88, 1.12)		2.7
61 61	-0.751	0.182	0.68 (0.78, 1.22)		-3.1	0.76 (0.86, 1.14)		-3.7
73 73	0.707	0.184	0.74 (0.78, 1.22)		-2.5	0.79 (0.85, 1.15)		-3.0
95 95	0.380	0.178	1.29 (0.78, 1.22)		2.4	1.19 (0.87, 1.13)		2.7

Bảng 3.7: Các câu hỏi không phù hợp với mô hình của Đề 1

3.1.4. Phân loại và chỉnh sửa câu hỏi

Để tăng chất lượng câu hỏi, nhóm tác giả sử dụng kết quả phân tích của Conquest với từng câu hỏi. Tất cả 522 câu hỏi phù hợp với mô hình được phân chia thành hai nhóm: nhóm câu hỏi tốt và nhóm câu hỏi cần chỉnh sửa.

Nhóm câu hỏi tốt phù hợp với mô hình Rasch, có độ khó và độ phân biệt chấp nhận được theo lý thuyết khảo thí cổ điển cùng với các phương án gây nhiễu đạt yêu cầu. Bảng 3.8 là kết quả phân tích của một câu hỏi tốt của Đề 4.

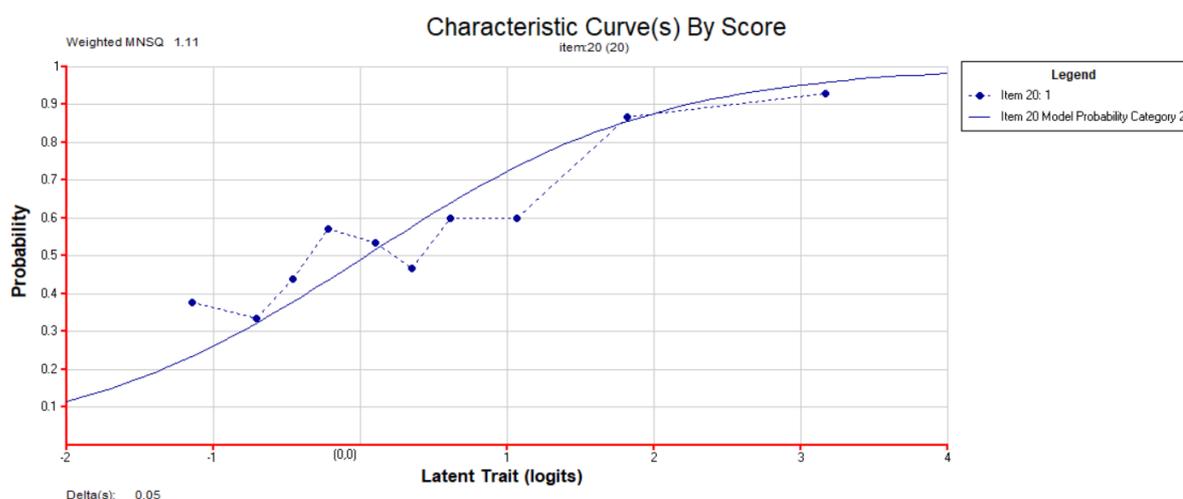
Bảng 3.8: Kết quả phân tích của câu hỏi 20 Đề 4

```

Item 20
-----
item:20 (20)
Cases for this item      150  Discrimination  0.34
Item Threshold(s):      0.05  Weighted MNSQ  1.11
Item Delta(s):          0.05
-----
Label   Score   Count   % of tot  Pt Bis   t (p)   WLEAvg:1  WLE SD:1
-----
a       0.00     26     17.33    -0.15   -1.84(.068)  0.05     0.95
b       1.00     85     56.67     0.34    4.36(.000)  0.90     1.52
c       0.00     28     18.67    -0.26   -3.25(.001) -0.17     0.83
d       0.00     11      7.33    -0.04   -0.46(.647)  0.19     0.61
=====

```

Với lý thuyết khảo thí cổ điển, độ khó của câu hỏi được tính là 0,57, nằm trong khoảng 0,25-0,75, cụ thể 56,67% thí sinh trả lời đúng câu hỏi này. Độ phân biệt rất tốt, $d = 0,34$, cho thấy câu hỏi có thể phân biệt giữa nhóm thí sinh có trình độ cao và nhóm thí sinh có năng lực thấp. Hệ số tương quan (PT BIS) cho chúng ta thấy những phương án gây nhiễu có chỉ số âm và phương án đúng có chỉ số dương. Các phương án đưa ra ở Câu 20 có giá trị đánh giá năng lực của thí sinh.



Hình 3.3: Đường cong đặc trưng của câu hỏi 20 – Đề 4

Kết quả này được khẳng định thông qua đường cong đặc trưng của câu hỏi. Nhìn Hình 3.3 ta có thể thấy đường cong đặc trưng của câu hỏi (đường nét đứt) bám sát đường cong kỳ vọng của câu hỏi, xuất phát từ điểm xác suất gần gốc tọa độ đi lên, cho thấy câu hỏi thiết kế phù hợp để kiểm tra năng lực của thí sinh.

Nhóm câu hỏi cần chỉnh sửa gồm các câu hỏi phù hợp với mô hình Rasch, có độ khó và độ phân biệt chấp nhận được theo lý thuyết khảo thí cổ điển. Tuy nhiên, nhà nghiên cứu cần đầu tư thời gian và công sức để chỉnh sửa các phương án nhiễu. Tổng hợp kết quả của bảy đề, số lượng câu hỏi cần chỉnh sửa là 84 câu hỏi.

Bảng 3.9: Tổng hợp số câu hỏi cần chỉnh sửa

Đề	1	2	3	4	5	6	7
Số câu hỏi cần chỉnh sửa	14	18	12	12	12	14	2

ConQuest: Generalised Item Response Modelling Software Tue Feb 14 12:16 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT			
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T	
1 1	-2.312	0.239	1.11 (0.79, 1.21)	1.0	0.92	(0.73, 1.27)	-0.5		
2 2	-2.492	0.250	1.54 (0.79, 1.21)	4.3	1.06	(0.70, 1.30)	0.4		
3 3	-2.256	0.236	0.82 (0.79, 1.21)	-1.7	0.95	(0.73, 1.27)	-0.3		
4 4	-3.914	0.404	0.98 (0.79, 1.21)	-0.1	0.92	(0.35, 1.65)	-0.1		
5 5	-0.893	0.189	1.27 (0.79, 1.21)	2.3	1.06	(0.84, 1.16)	0.7		
...									
94 94	-0.751	0.188	1.24 (0.79, 1.21)	2.1	1.19	(0.84, 1.16)	2.1		
95 95	-3.622	0.362	0.70 (0.79, 1.21)	-3.0	0.95	(0.44, 1.56)	-0.1		
96 96	-3.383	0.332	0.86 (0.79, 1.21)	-1.4	0.99	(0.51, 1.49)	0.0		
97 97	-1.995	0.222	1.64 (0.79, 1.21)	5.0	1.16	(0.77, 1.23)	1.3		
98 98	-1.462	0.202	0.98 (0.79, 1.21)	-0.1	1.06	(0.82, 1.18)	0.7		
99 99	-0.337	0.185	0.83 (0.79, 1.21)	-1.6	0.92	(0.84, 1.16)	-0.9		
100 100	0.004	0.185	1.74 (0.79, 1.21)	5.6	1.47	(0.83, 1.17)	4.8		

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.960

Chi-square test of parameter equality = 5947.34, df = 100, Sig Level = 0.000

[^] Quick standard errors have been used

=====

Hình 3.4: Kết quả phân tích Conquest của Đề 6

Phần này lấy ví dụ chỉnh sửa các câu hỏi ở Đề 6. Sau khi phân tích với Conquest, Đề 6 loại chín câu hỏi không phù hợp với mô hình phân tích. Các câu hỏi còn lại (91 câu) phù hợp với mô hình phân tích của Đề 6 được chia thành hai nhóm: nhóm (1) là 77 câu hỏi tốt đảm bảo chất lượng, các chỉ số tương đương câu hỏi 20 của Đề 4 được phân tích ở trên; nhóm (2) gồm 14 câu hỏi phải chỉnh sửa là những câu hỏi có các phương án nhiễu chưa đạt yêu cầu. Các câu hỏi sau khi chỉnh sửa được chuyển cho các chuyên gia để thẩm định trước các bước tiếp theo của nghiên cứu. Bảng 3.10 trình bày kết quả chỉnh sửa các câu hỏi của đề 6.

Bảng 3.10: Các câu hỏi cần chỉnh sửa của Đề 6

STT	Mã câu hỏi	Nội dung câu hỏi	Các lựa chọn	Đáp án	Phương án nhiễu cần chỉnh sửa	Kết quả chỉnh sửa và thẩm định bởi chuyên gia
1	D6:61 L3	mail: Do they have any "mail"?	a. nơi ở b. tiền bạc c. thư từ d. quần áo ấm	c	b	a. nơi ở b. đồ đạc c. thư từ d. quần áo ấm
2	D6:50 L3	slowly: They worked "slowly".	a. dễ dàng b. chậm chạp c. một mình d. hằng ngày	b	d	a. một cách dễ dàng b. một cách chậm chạp c. một mình d. một cách cẩn trọng
3	D6:11 L1	minute: It takes me some "minutes" to go to school.	a. hành động b. tiền c. chặng đường d. phút	d	a	a. nỗ lực b. tiền c. chặng đường d. phút
4	D6:7L 1	room: Where is the "room"?	a. quyển sách b. xe ô tô c. cửa hàng d. phòng	d	b	a. quyển sách b. hòn đá c. cửa hàng d. phòng

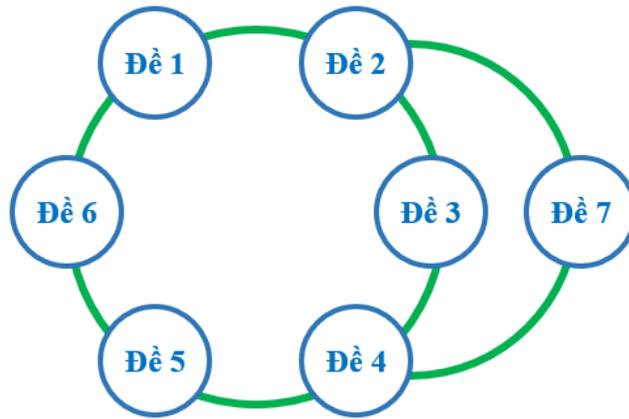
STT	Mã câu hỏi	Nội dung câu hỏi	Các lựa chọn	Đáp án	Phương án nhiều cần chỉnh sửa	Kết quả chỉnh sửa và thẩm định bởi chuyên gia
5	D6:18 L2	bank: I like this "bank".	a. khoảng thời gian b. nhóm công nhân c. ngân hàng d. thành phần	c	bd	a. khoảng thời gian b. không gian c. ngân hàng d. yếu tố
6	D6:14 L1	health: "Health" is important.	a. học tập b. sức khỏe c. luyện tập d. nhận sự giúp đỡ	b	d	a. học tập b. sức khỏe c. luyện tập d. sự giúp đỡ
7	D6:25 L2	clean: It was "clean".	a. nguy hiểm b. sạch sẽ c. rất tốt d. rất nhỏ	b	d	a. nguy hiểm b. sạch sẽ c. rất tốt d. lấp lánh
8	D6:4L 1	different: They are "different".	a. dễ thấy b. to lớn c. rất khó d. khác biệt	d	a	a. dễ dãi b. to lớn c. rất khó d. khác biệt
9	D6:19 L2	surprise: It is a big "surprise".	a. sức mạnh để thay đổi một cái gì đó b. lễ hội quốc tế c. bất ngờ d. quá trình làm một cái gì đó	c	d	a. sức mạnh b. lễ hội c. bất ngờ d. giải thưởng
10	D6:8L 1	share: We "shared" an office for years.	a. đến nơi b. tìm kiếm c. theo dõi d. chia sẻ	d	a	a. xây dựng b. tìm kiếm c. theo dõi d. chia sẻ
11	D6:91 L5	grammar: We like "grammar".	a. ghi chép chi tiêu b. hợp tác c. ngữ pháp d. sửa chữa	c	d	a. ghi chép chi tiêu b. sự hợp tác c. ngữ pháp d. việc sửa chữa

STT	Mã câu hỏi	Nội dung câu hỏi	Các lựa chọn	Đáp án	Phương án nhiều cần chỉnh sửa	Kết quả chỉnh sửa và thẩm định bởi chuyên gia
12	D6:63 L4	impose: They "imposed" their ideas.	a. viết ra b. giữ vững c. trình bày d. áp đặt	d	b	a. viết ra b. phát triển c. trình bày d. áp đặt
13	D6:26 L2	debate: We had a "debate".	a. bài kiểm tra b. đề cử lãnh đạo c. kết quả tậ d. cuộc tranh luận	d	a	a. cuộc hẹn b. đề cử lãnh đạo c. kết quả tậ d. cuộc tranh luận
14	D6:36 L3	confirm: They "confirmed" it.	a. nghi ngờ b. thử nghiệm c. đòi hỏi d. xác nhận	d	c	a. nghi ngờ b. thử nghiệm c. tạo dựng d. xác nhận

3.1.5. Cân bằng đề và chuẩn hóa ngân hàng câu hỏi

Sau khi phân tích chất lượng đề và câu hỏi của các đề thử nghiệm, các câu hỏi không phù hợp với mô hình sẽ bị loại bỏ, nhà nghiên cứu tiến hành cân bằng đề với phần mềm R, gói equateIRT, để đưa các tham số câu hỏi về cùng một thang đo của ngân hàng câu hỏi trắc nghiệm thích ứng.

Cần lưu ý rằng trong nghiên cứu này, các đề được thiết kế có số câu hỏi neo như trong Bảng 3.11 và Hình 3.5 trước khi đưa vào thử nghiệm để chuẩn bị cho bước cân bằng đề này. Có thể lấy Đề 1 làm ví dụ minh họa: Đề 1 có thiết kế neo với Đề 2 và Đề 6. Sau khi phân tích Conquest, Đề 1 có 91 câu hỏi được giữ lại, trong đó có 10 câu hỏi neo với Đề 2 và 13 câu hỏi neo với Đề 6.



Hình 3.5: Sơ đồ neo giữa các đề

Bảng 3.11: Thiết kế câu hỏi neo giữa bảy đề sau khi phân tích Conquest

Đề	1	2	3	4	5	6	7
1	90	10				13	
2		91	13				4
3			88	14			
4				92	12		22
5					86	11	
6						91	
7							83

Số câu hỏi neo giữa 2 đề

Số câu hỏi phù hợp mô hình sau khi phân tích Conquest

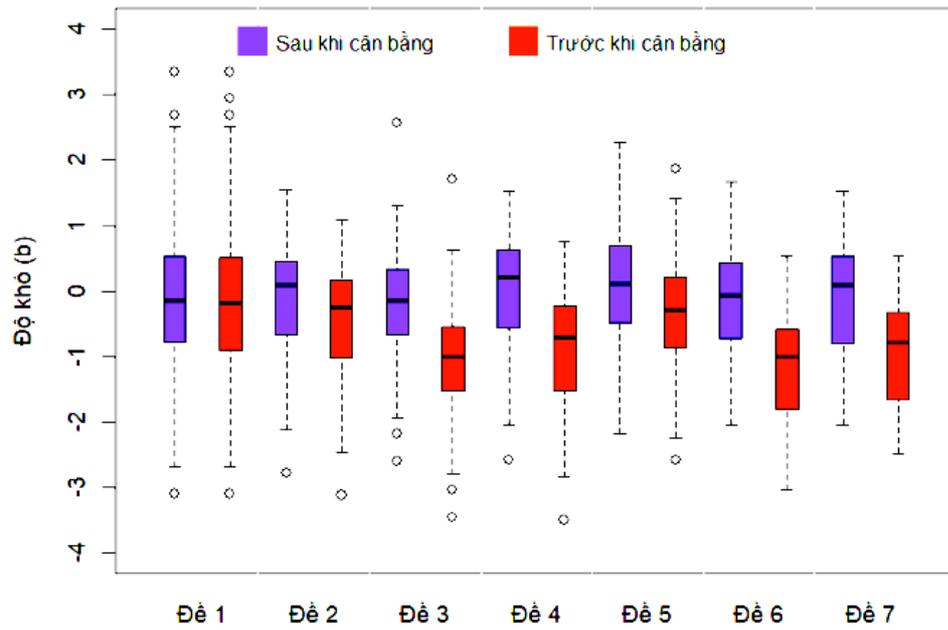
Sử dụng phần mềm thống kê R với phương pháp tính của Loyd và Hoover (1980), các hệ số cân bằng của bảy đề thi (theo lược đồ neo) được tính toán và trình bày trong Bảng 3.12.

Bảng 3.12: Hệ số cân bằng bảy đề thử nghiệm

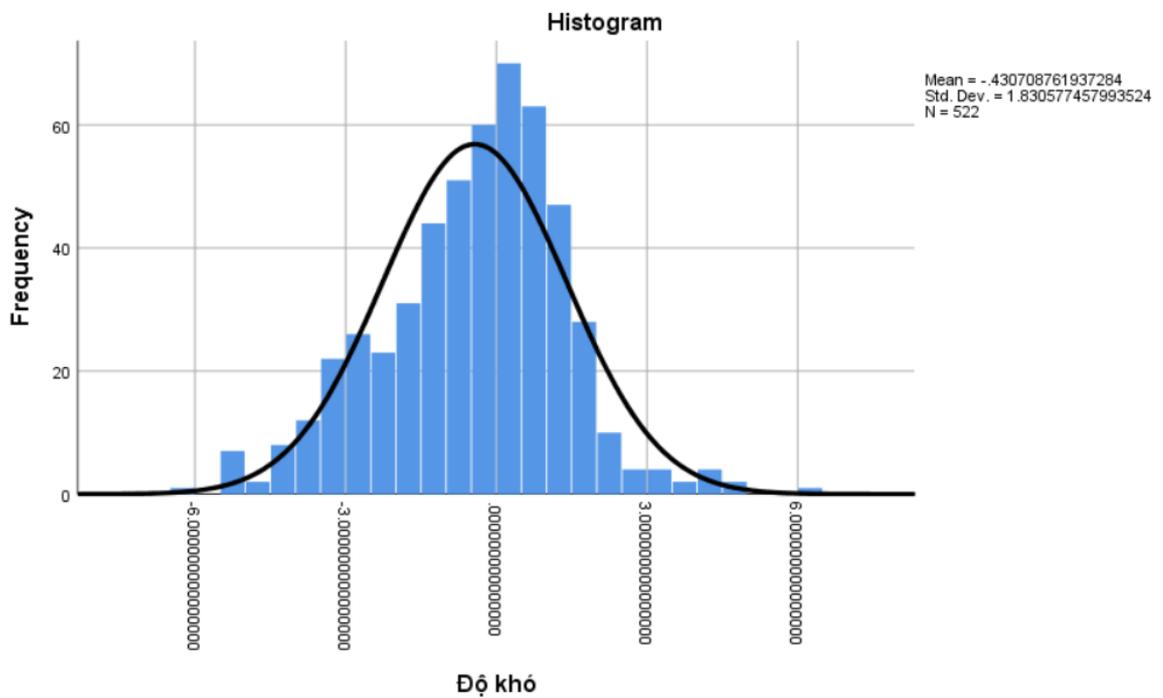
Đề	Hệ số	Giá trị	Sai số	Hệ số	Giá trị	Sai số
1	A	1.000	0.000	B	0.000	0.000
2	A	1.000	0.000	B	0.343	0.093
3	A	1.000	0.000	B	0.853	0.094
4	A	1.000	0.000	B	0.906	0.100
5	A	1.000	0.000	B	0.398	0.092
6	A	1.000	0.000	B	0.994	0.084
7	A	1.000	0.000	B	0.860	0.105

Tham số độ khó của câu hỏi được ước lượng bằng mô hình Rasch 1 tham số (1PL – model), do vậy các hệ số cân bằng $A_{s,s+1}$ đều bằng 1. Các hệ số cân bằng $B_{s,s+1}$ tương ứng của 7 đề được trình bày trong bảng 3.12. Tiếp tục sử dụng các công thức của Loyd và Hoover (1980), độ khó của câu hỏi trong bảy đề được quy về cùng một thang đo. Độ khó của các câu hỏi sử dụng làm câu hỏi neo sau khi cân bằng vẫn đảm bảo khoảng cách tương đối về độ khó. Hơn nữa, sau khi cân bằng độ khó của các đề thi, độ khó của câu hỏi neo thuộc hai đề thi khác nhau đã được quy về một độ khó duy nhất. Các câu hỏi khác trong đề thi cũng được cân bằng theo công thức của Loyd và Hoover (1980). Biểu đồ 3.6 so sánh độ khó của các câu hỏi thuộc bảy đề trước và sau khi cân bằng.

Từ biểu đồ cho thấy, các đề thi thử nghiệm được xây dựng dựa trên cùng một bản đặc tả đề thi (cấu trúc theo nội dung và các cấp độ năng lực) nhưng chưa đảm bảo điều kiện để các đề thi có sự tương đương về độ khó. Phổ độ khó của đề thi sau khi cân bằng có tính ổn định hơn phổ độ khó của các đề thi khi chưa cân bằng (Hình 3.6) nên các câu hỏi này đảm bảo đủ tính khoa học và tính tương thích để đưa vào xây dựng một ngân hàng câu hỏi có số câu hỏi lớn.



Hình 3.6: Độ khó câu hỏi thi trước và sau khi cân bằng



Hình 3.7: Độ khó của ngân hàng câu hỏi chuẩn hóa

Sau quá trình cân bằng đề, độ khó của ngân hàng câu hỏi được định cỡ trong khoảng -6 đến 6, các câu hỏi này được đưa vào ngân hàng câu hỏi theo 5 bậc thông dụng của danh sách từ NGSL cùng với các thông số về độ khó. Từ ngân hàng này, hệ thống trắc nghiệm thích ứng sẽ lựa chọn câu hỏi theo thuật toán thích ứng định trước để khởi tạo các bài trắc nghiệm cho thí sinh.

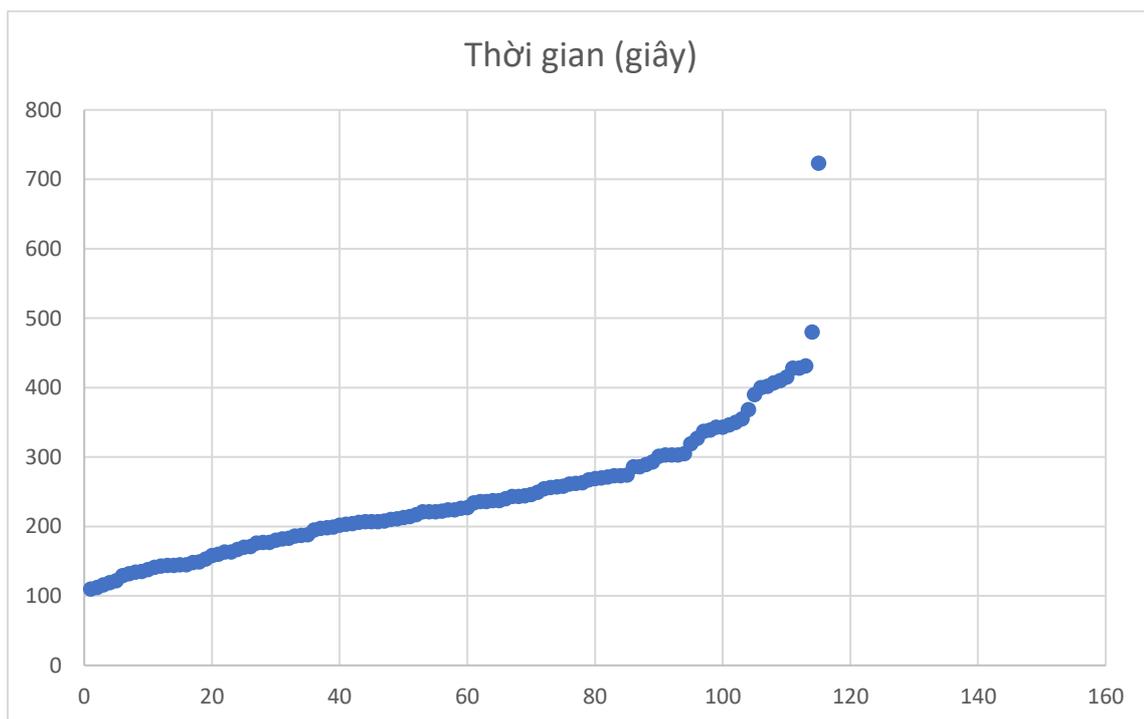
Độ khó của câu hỏi trong ngân hàng đề có phân phối chuẩn đạt yêu cầu của hệ thống trắc nghiệm thích ứng, tuy nhiên cũng mở ra những thách thức tiếp theo khi số lượng các câu hỏi có độ khó trên 2 còn hạn chế trong ngân hàng câu hỏi vừa được xây dựng. Việc biên soạn và thử nghiệm các câu hỏi để bổ sung vào ngân hàng câu hỏi cần được ưu tiên trong kế hoạch hoàn thiện hệ thống trắc nghiệm thích ứng.

3.2. Đánh giá bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh

3.2.1. Quá trình làm bài của thí sinh trên hệ thống UEd-CAT

Thời gian làm bài

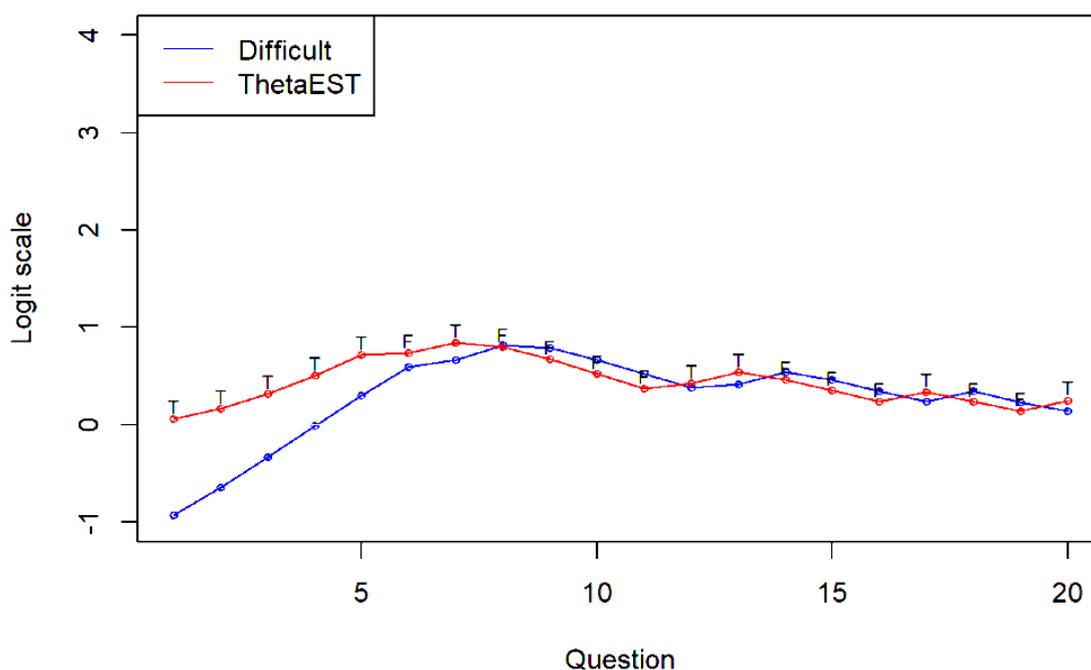
Thời gian làm bài giao động từ 12 giây với 6 câu hỏi đến 12 phút 3 giây (723 giây) với 20 câu hỏi. Với đề dài nhất (20 câu hỏi), thời gian làm bài dao động từ 110 giây đến 723 giây. Có thể thấy hiệu quả thời gian thể hiện rõ rệt khi số lượng câu hỏi trong bài thi giảm đi nhiều so với đề cố định truyền thống với 100 câu hỏi.



Hình 3.8: Thời gian làm bài trên hệ thống UEd-CAT

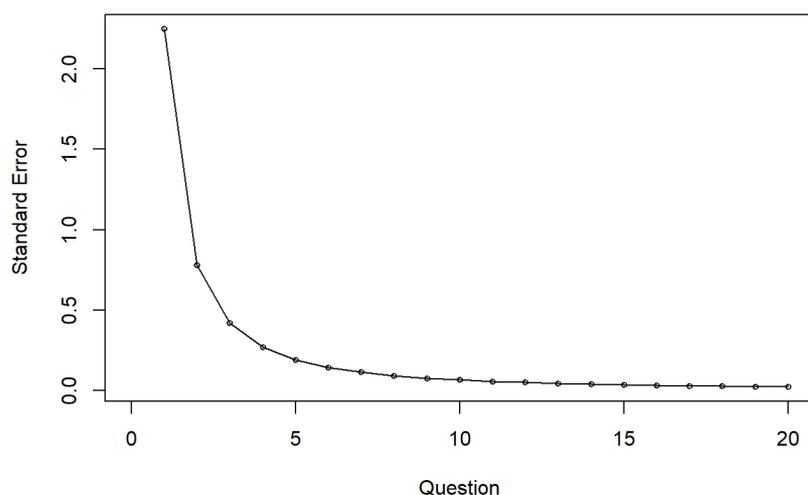
Lộ trình thích ứng độ khó câu hỏi trên hệ thống

Hình 3.9 minh họa lộ trình thích ứng với độ khó câu hỏi được thực hiện vào ngày 27 tháng 5 bởi một thí sinh có được mã hóa là HONG. Tổng số câu hỏi trong bài kiểm tra là 20. Thí sinh đã đưa ra 10 câu trả lời đúng và đạt số điểm 59,93. Có thể nhận thấy rằng khi HONG trả lời đúng một câu thì độ khó của câu tiếp theo sẽ cao hơn và khi HONG trả lời sai câu hỏi thì hệ thống sẽ phản hồi theo hai hướng để phù hợp với khả năng của thí sinh. Với câu trả lời sai ở câu hỏi số 8, hệ thống cung cấp cho thí sinh một câu tiếp theo với độ khó thấp hơn. Tuy nhiên, với câu trả lời sai ở câu hỏi số 6, hệ thống vẫn đưa ra câu hỏi tiếp theo với độ khó cao hơn nhưng sự chênh lệch về độ khó đã giảm đi so với sự chênh lệch độ khó giữa câu hỏi số 4 và câu hỏi số 5.



Hình 3.9: Lộ trình thích ứng trong bài kiểm tra của thí sinh HONG

Quá trình ước lượng năng lực được lặp đi lặp lại sau mỗi câu hỏi trên cơ sở câu trả lời của thí sinh và độ khó của câu hỏi đó. Sai số chuẩn của phép ước lượng giảm dần trong quá trình cho đến khi thỏa mãn tiêu chí dừng của sai số chuẩn được xác định trước ở mức 0,02 trong hệ thống UEd-CAT. Trong trường hợp này, bài thi kết thúc sau khi thí sinh trả lời câu hỏi số 20 và hệ thống kết thúc ước tính năng lực của thí sinh với sai số chuẩn là 0,0227 (Hình 3.10).



Hình 3.10: Sai số chuẩn của phép ước lượng năng lực cập nhật sau từng câu hỏi trong bài làm của thí sinh HONG

Mức độ đa dạng câu hỏi và gói câu hỏi trên hệ thống

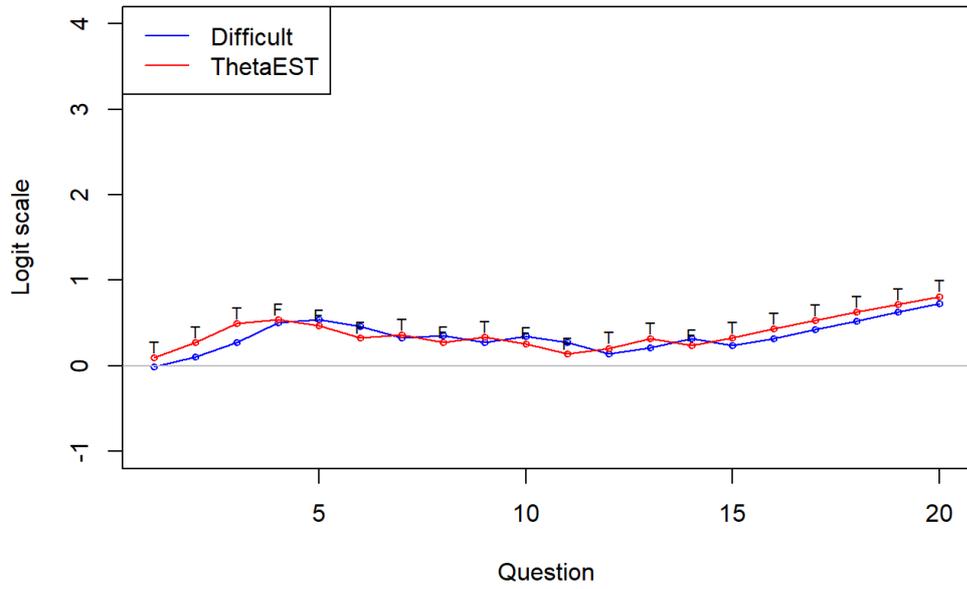
Bảng 3.13: Các gói câu hỏi

Số câu hỏi mỗi bài	6	8	10	11	12	13	14	15	16	17	18	19	20	Tổng
Số bài	2	2	3	4	20	5	2	4	23	10	2	17	115	209

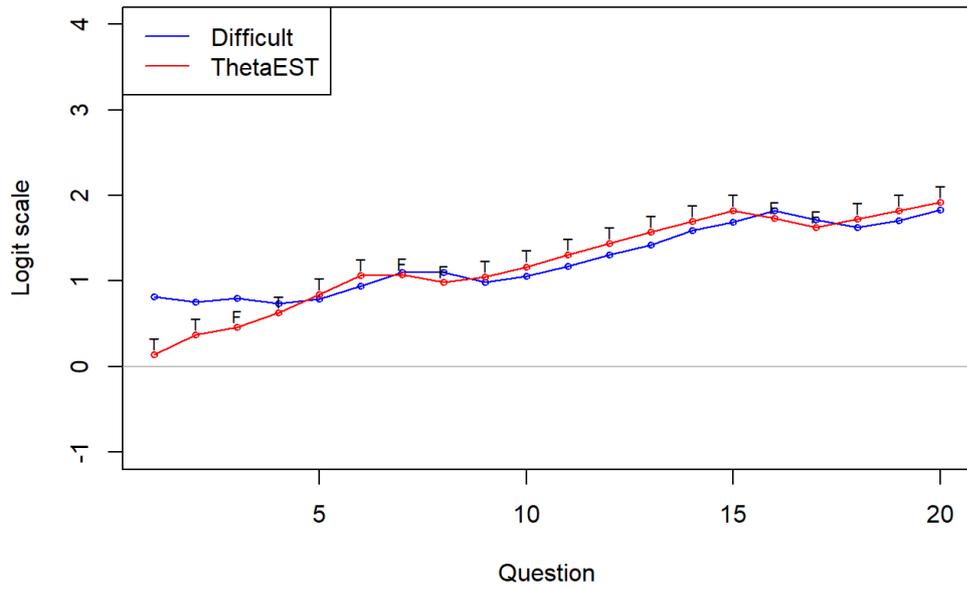
Hệ thống cung cấp các gói câu hỏi khác nhau, từ 6 đến 20 câu hỏi mỗi bài thi cho thí sinh tham gia. Với cỡ mẫu 209 bài thi thí sinh thực hiện, gói 20 câu hỏi có tần suất cao nhất, có 115 lượt làm bài thí sinh trả lời 20 câu hỏi, các gói 6,8,14 và 18 câu hỏi có tần suất thấp nhất.

Bảng 3.14: Ba lượt làm bài của thí sinh DAN

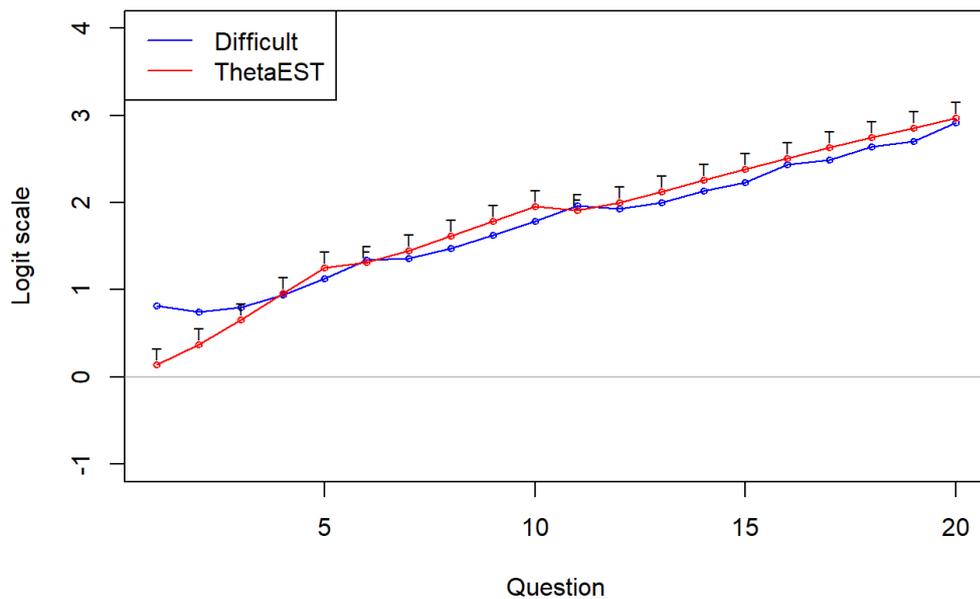
Thí sinh: DAN								
Ngày thực hiện: 22.5			Ngày thực hiện: 27.5			Ngày thực hiện: 6.6		
STT	Mã câu hỏi	Độ khó	STT	Mã câu hỏi	Độ khó	STT	Mã câu hỏi	Độ khó
1	D5:25L2	-0.0082	1	D7:88L5	0.82	1	D7:88L5	0.82
2	D1:81L4	0.1089	2	D1:78L4	0.7519	2	D3:30L2	0.7498
3	D5:34L3	0.2725	3	D1:46L3	0.7988	3	D1:46L3	0.7988
4	D7:63L4	0.5062	4	D3:62L3	0.7409	4	D1:82L5	0.9431
5	D3:91L5	0.5458	5	D3:69L4	0.7881	5	D2:78L4	1.1326
6	D4:36L3	0.4619	6	D1:82L5	0.9431	6	D7:93L5	1.344
7	D5:55L3	0.3249	7	D1:23L2	1.1019	7	D4:31L2	1.3587
8	D1:94L5	0.3541	8	D6:57L3	1.1016	8	D7:87L5	1.4732
9	D7:76L4	0.2739	9	D2:89L5	0.9874	9	D4:85L5	1.63
10	D2:13L1	0.3453	10	D5:71L4	1.0546	10	D2:83L5	1.7891
11	D5:32L3	0.2735	11	D7:86L5	1.17	11	D2:95L5	1.9612
12	D4:60L3	0.1414	12	D6:74L4	1.3071	12	D4:74L4	1.9286
13	D6:98L5	0.2154	13	D5:96L5	1.4183	13	D7:55L3	2.0014
14	D1:52L3	0.3178	14	D4:99L5	1.5889	14	D7:95L5	2.1368
15	D1:37L3	0.2361	15	D7:73L4	1.6915	15	D4:47L3	2.2349
16	D3:56L3	0.3203	16	D4:52L3	1.8215	16	D2:72L4	2.4374
17	D5:29L2	0.4287	17	D5:78L4	1.7157	17	D2:100L5	2.488
18	D7:54L3	0.5278	18	D4:85L5	1.63	18	D5:95L5	2.6408
19	D3:75L4	0.6265	19	D2:73L4	1.7065	19	D1:35L3	2.7066
20	D4:67L4	0.7274	20	D2:71L4	1.8304	20	D6:38L3	2.9185



Hình 3.11: Lộ trình thích ứng lượt làm bài số 1 của thí sinh DAN



Hình 3.12: Lộ trình thích ứng lượt làm bài số 2 của thí sinh DAN



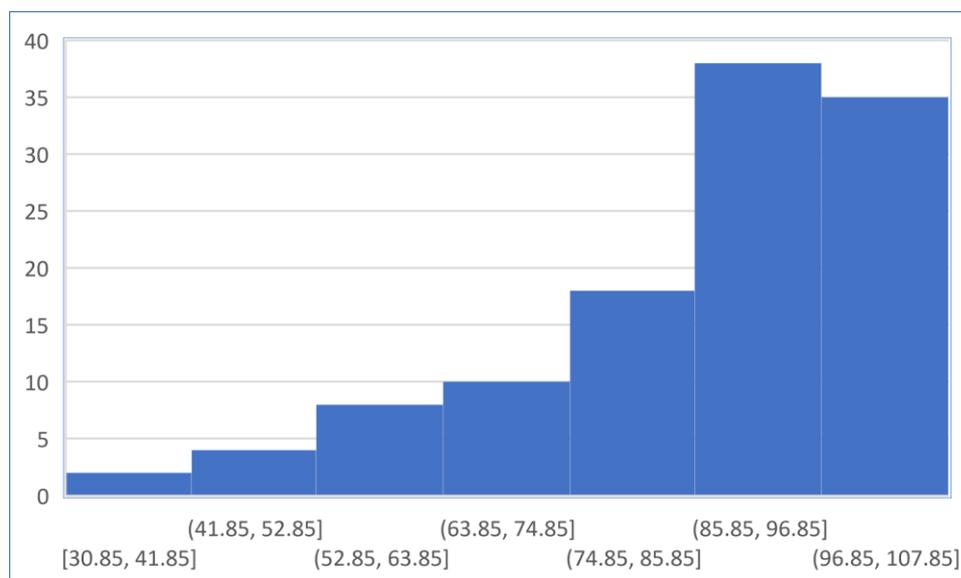
Hình 3.13: Lộ trình thích ứng lượt làm bài số 3 của thí sinh DAN

Đáng chú ý là mặc dù một người làm bài kiểm tra có thể có cùng số lượng câu hỏi trong các lượt làm bài, nhưng người đó sẽ có những câu hỏi khác nhau cũng như những lộ trình thích ứng khác nhau về độ khó câu hỏi trong mỗi lượt làm bài. Tỷ lệ các câu hỏi trùng lặp giữa các bài kiểm tra khác nhau được giảm thiểu nhờ một ngân hàng câu hỏi đủ lớn. Một ví dụ là trường hợp một thí sinh đã thực hiện ba lượt làm bài trên hệ thống sẽ được sử dụng để minh họa mức tăng khác nhau về độ khó của câu hỏi trong gói 20 câu và mức độ trùng lặp câu hỏi giữa các lượt làm bài. Có thể thấy từ Bảng 3.14 rằng trong số 60 câu hỏi thí sinh DAN trả lời trên hệ thống, chỉ có bốn câu hỏi được lặp lại trong lượt kiểm tra thứ hai và thứ ba, bao gồm các câu hỏi có mã D7:88L5, D1:46L3, D1:82L5 và D4:85L5.

3.2.2. Kết quả làm bài của thí sinh trên hệ thống

Trong phần này, kết quả làm bài trên hệ thống của các thí sinh tham gia thử nghiệm sẽ được phân tích để khám phá độ giá trị và độ tin cậy của công cụ trong việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh.

Khả năng phân biệt của các bài kiểm tra có cùng số lượng câu hỏi



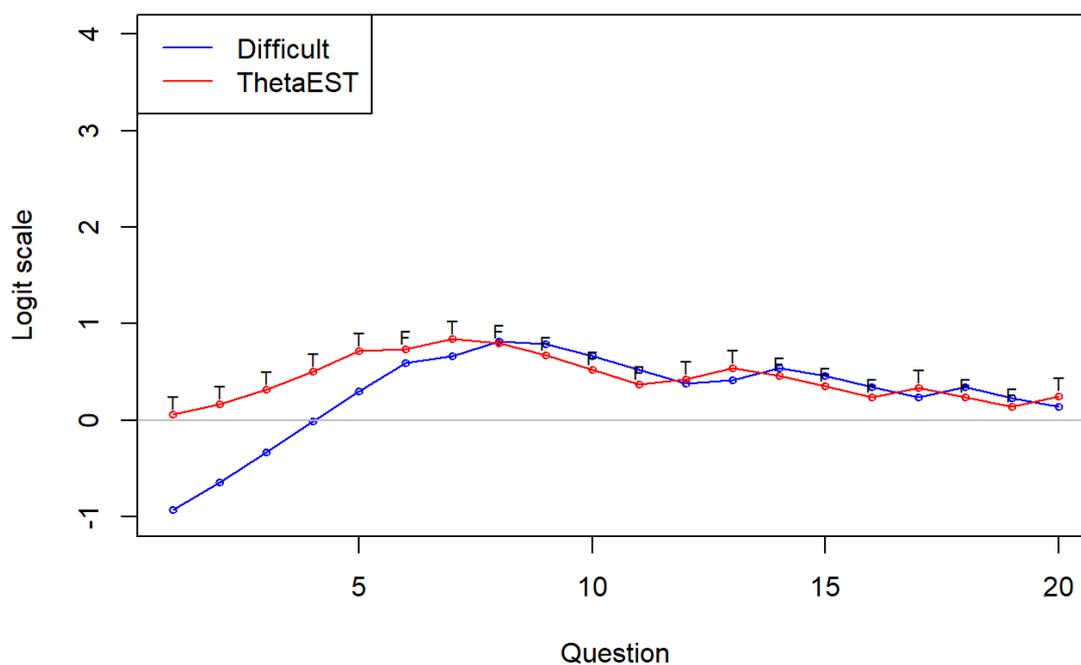
Hình 3.14: Điểm của thí sinh với bài kiểm tra 20 câu hỏi

Với cùng số lượng câu hỏi trong bài thi, kết quả bài thi sẽ khác nhau phản ánh các mức độ khác nhau của thí sinh về từ vựng tiếp nhận tiếng Anh. Như có thể thấy trên Hình 3.14, các bài kiểm tra khác nhau với cùng số lượng câu hỏi (20 câu) có thể chia thí sinh thành các nhóm năng lực khác nhau: “trình độ thấp” với số điểm dưới 52,85; “trình độ trung bình” với số điểm lên tới 85,85; và “trình độ cao” với số điểm trên 85,85. Đây là dấu hiệu cho thấy bài kiểm tra có thể phân loại các nhóm thí sinh khác nhau theo năng lực của họ dù có cùng số câu hỏi trong bài kiểm tra.

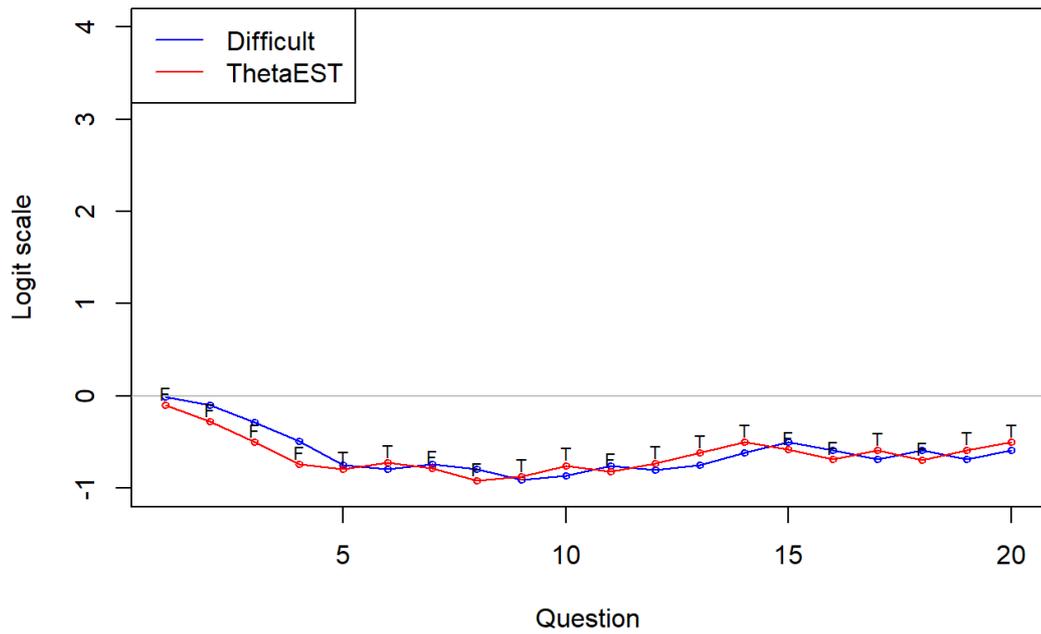
Cùng bài kiểm tra 20 câu hỏi, cùng tỉ lệ trả lời đúng là 50%, kết quả của các thí sinh cũng cho thấy sự phân loại rõ rệt, cho thấy sự đa dạng câu hỏi và gói câu hỏi, cũng như khả năng phân loại của hệ thống dựa trên năng lực làm bài của thí sinh và độ khó của câu hỏi.

Bảng 3.15: Thông số của các bài kiểm tra 20 câu hỏi ba thí sinh thực hiện

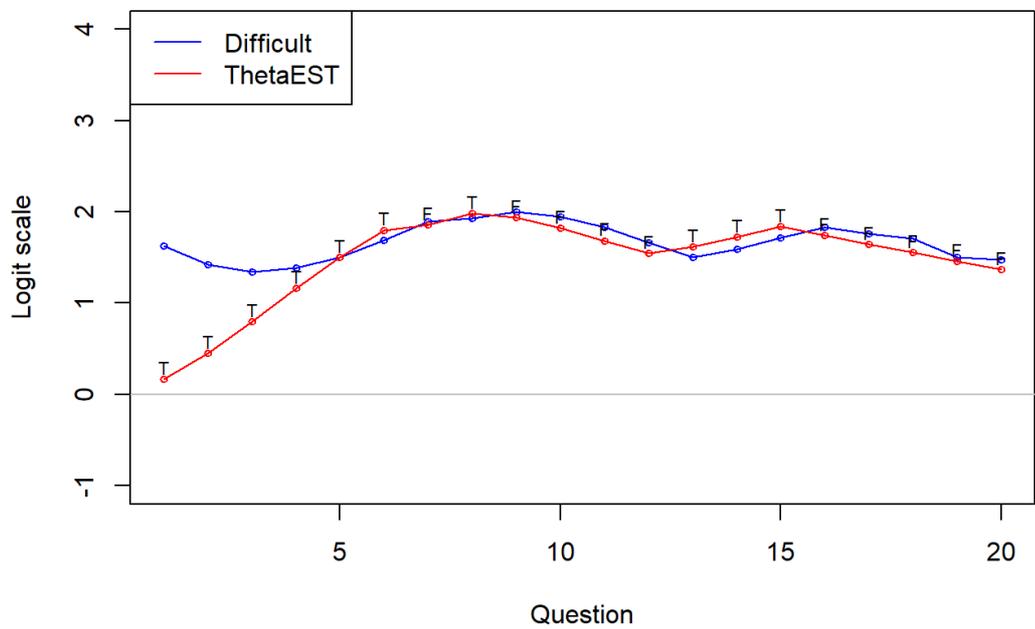
Thí sinh	HONG	MDUC	LINH
Thời gian làm bài	2023-05-27	2023-05-29	2023-05-31
Số câu hỏi trong bài	20	20	20
Số câu trả lời đúng	10	10	10
Kết quả	59.93	30.85	91.49
Sai số chuẩn	0.0227	0.0224	0.0228



Hình 3.15: Lộ trình thích ứng trong bài trắc nghiệm HONG thực hiện



Hình 3.16: Lộ trình thích ứng trong bài trắc nghiệm MDUC thực hiện



Hình 3.17: Lộ trình thích ứng trong bài trắc nghiệm LINH thực hiện

Sự nhất quán của kết quả làm bài của một thí sinh trong các bài kiểm tra khác nhau

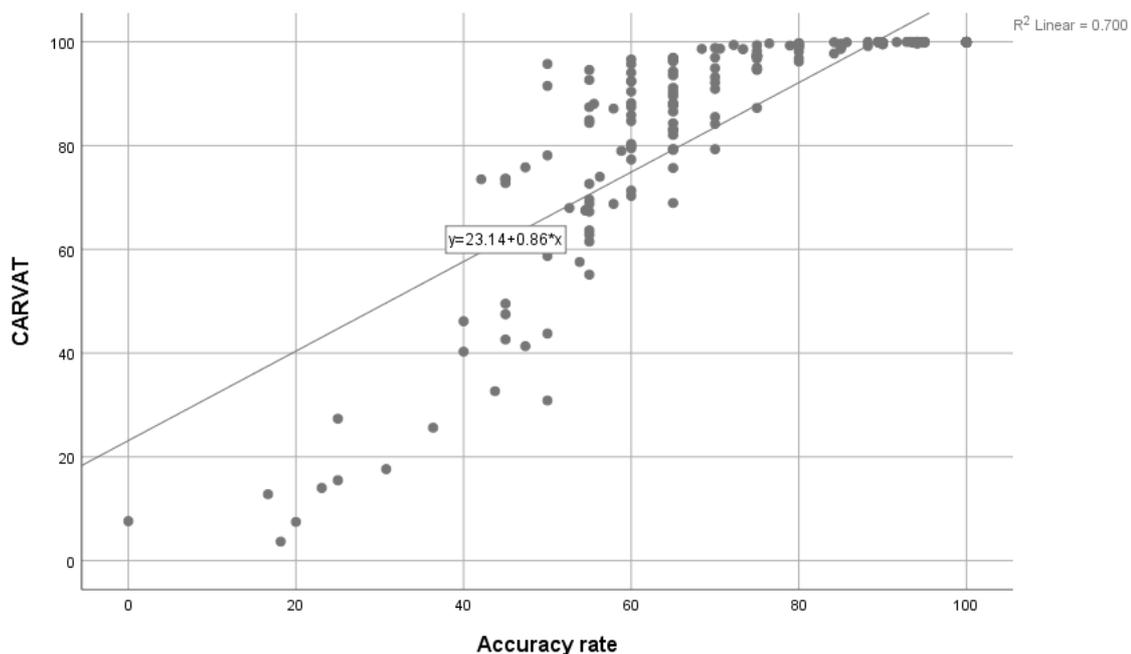
Bảng 3.16 thể hiện kết quả làm bài trên hệ thống của thí sinh DANH và TDUC. Trong khi DANH thực hiện ba bài kiểm tra 20 câu hỏi thì TDUC thực hiện ba bài kiểm tra với số lượng câu hỏi khác nhau. Dù có sự khác biệt này, kết quả làm bài họ nhận được vẫn đủ nhất quán để thể hiện năng lực cao của họ.

Bảng 3.16: Kết quả làm bài trong các lượt làm bài khác nhau

Thí sinh	TDUC			DANH		
	1	2	3	1	2	3
Bài kiểm tra trên hệ thống						
Số câu hỏi/bài kiểm tra	20	17	10	20	20	20
Số câu trả lời đúng	19	15	8	15	13	18
Kết quả làm bài kiểm tra	99.91	99.97	99.63	97.24	96.24	96.85

Tương quan giữa tỉ lệ trả lời chính xác và điểm kết quả làm bài kiểm tra thích ứng

Như có thể thấy từ Bảng 3.17, giá trị Sig. (2-Tailed) là 0,000, nhỏ hơn 0,05. Có thể kết luận là mối tương quan có ý nghĩa thống kê giữa hai biến nghiên cứu. Hệ số 0,837 trong phân tích tương quan Pearson cho thấy mối tương quan dương, rất mạnh giữa tỉ lệ trả lời chính xác của thí sinh và điểm của bài kiểm tra thích ứng. Có thể hiểu rằng tỉ lệ trả lời chính xác của thí sinh phản ánh năng lực của thí sinh trong bài kiểm tra thích ứng, từ đó cung cấp thêm bằng chứng về độ chính xác của trắc nghiệm thích ứng trong việc đánh giá năng lực thí sinh.



Hình 3.18: Biểu đồ phân tán tỉ lệ trả lời chính xác và điểm bài kiểm tra thích ứng

Bảng 3.17: Phân tích tương quan Pearson giữa tỉ lệ trả lời chính xác và điểm bài kiểm tra thích ứng

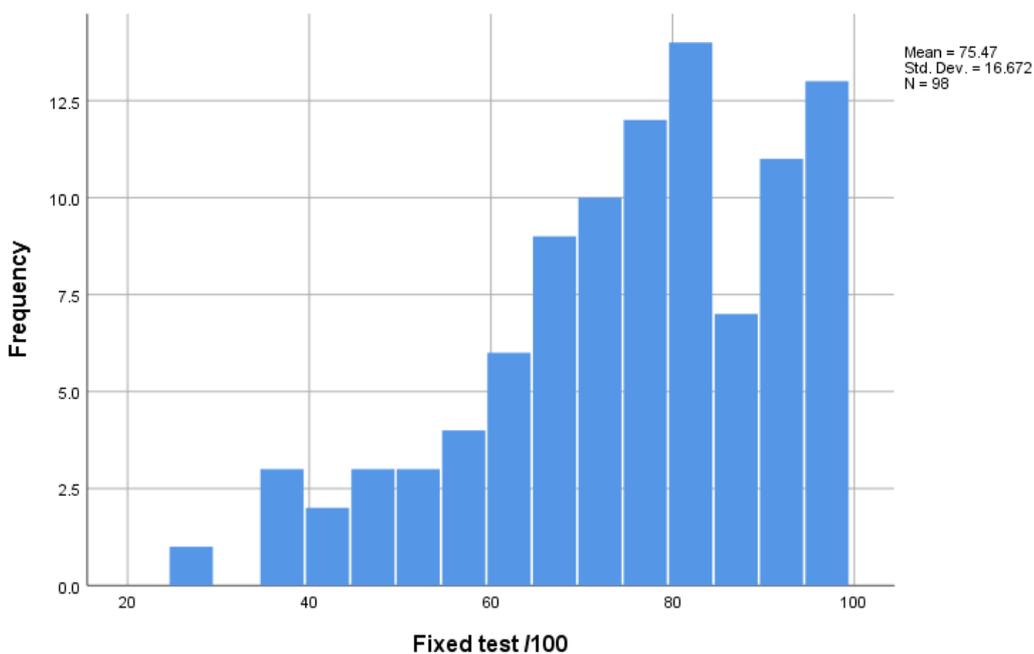
		Điểm bài kiểm tra thích ứng	Tỉ lệ trả lời chính xác
Điểm bài kiểm tra thích ứng	Pearson Correlation	1	.837**
	Sig. (2-tailed)		.000
	N	209	209
Tỉ lệ trả lời chính xác	Pearson Correlation	.837**	1
	Sig. (2-tailed)	.000	
	N	209	209
** . Correlation is significant at the 0.01 level (2-tailed).			

Tương quan giữa kết quả làm bài với bài kiểm tra cố định và bài kiểm tra thích ứng

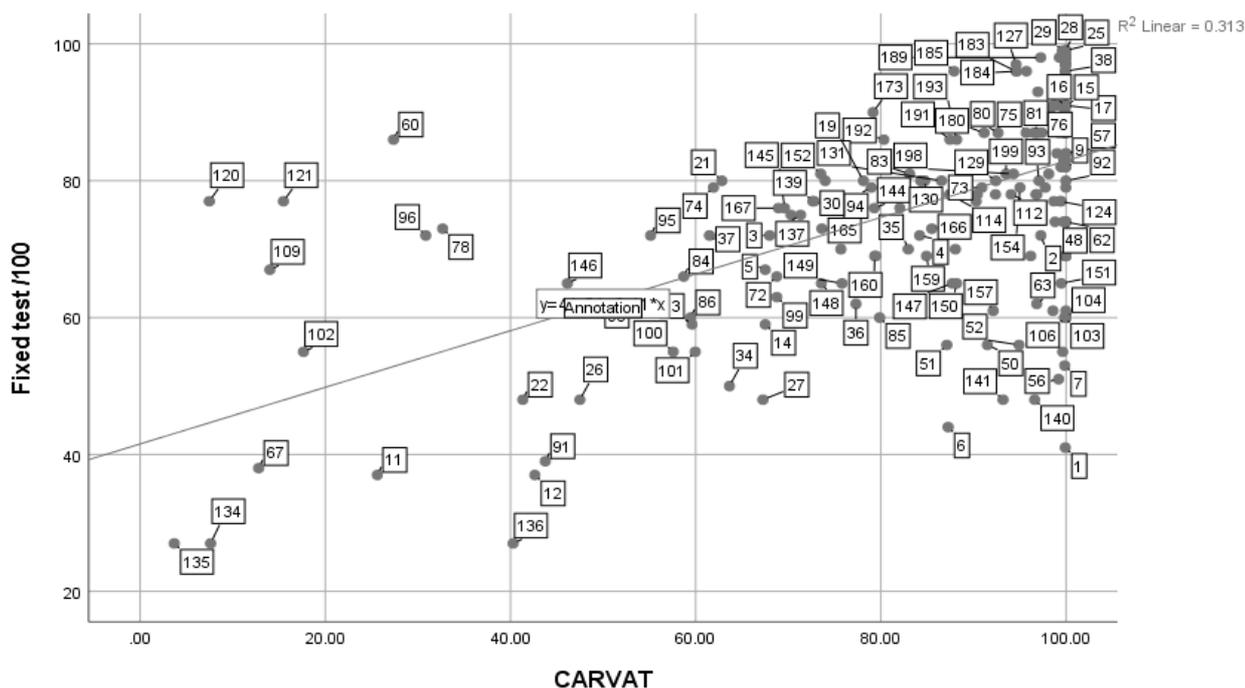
Trong phần này, phân tích tương quan được thực hiện giữa điểm bài kiểm tra cố định và bài kiểm tra thích ứng để cung cấp thêm bằng chứng xác thực về tính chính xác và độ tin cậy của bài kiểm tra thích ứng.

Kết quả làm bài kiểm tra cố định

Điểm kiểm tra của 98 người tham gia cả quá trình thử nghiệm trong bài kiểm tra cố định được tổng hợp để cung cấp cái nhìn tổng quan về trình độ từ vựng của mẫu nghiên cứu. Điểm của thí sinh dao động từ 27 đến 99. Điểm trung bình là 75,47 và 66,3% thí sinh đạt điểm từ 70 đến 99. Như có thể thấy từ biểu đồ, những người tham gia có nhiều cấp độ kiến thức từ vựng khác nhau, từ thấp đến cao và phần lớn thuộc về trình độ cấp trung bình và cao.



Hình 3.19: Kết quả làm bài kiểm tra cố định của 98 thí sinh



Hình 3.20: Biểu đồ phân tán điểm số trong bài kiểm tra cố định và bài kiểm tra thích ứng

Bảng 3.18: Các trường hợp bị loại trừ

Mã	Điểm bài kiểm tra cố định (/100)	Thời gian làm bài trắc nghiệm thích ứng (giây)	Điểm bài kiểm tra thích ứng (/100)	Số câu hỏi	Số câu trả lời đúng	Tỉ lệ trả lời đúng (%)
#121	77	21	15.48	8	2	25
#109	67	31	14.01	13	3	23
#120	77	32	7.47	10	2	20
#60	86	36	27.36	8	2	25

Cần lưu ý rằng mẫu được nghiên cứu là 209 vì mọi lượt làm bài của mỗi thí sinh trên hệ thống đều được tính vào cỡ mẫu. Hình 3.20 cho thấy có mối quan hệ tiềm ẩn giữa điểm số của bài kiểm tra cố định và bài kiểm tra thích ứng. Một số trường hợp quan sát được cho thấy đường đi bất thường trong biểu đồ phân tán. Sau khi kiểm tra lại thông tin về quá trình làm bài của các thí sinh, quyết định được đưa ra là loại trừ bốn trường hợp trong các phân

tích tiếp theo do thời gian thử nghiệm rất ngắn để tránh những tác động tiêu cực đến kết quả nghiên cứu.

Bảng 3.19: Phân tích tương quan Pearson giữa điểm bài kiểm tra cố định và điểm bài kiểm tra thích ứng

		Kết quả bài kiểm tra cố định	Kết quả bài kiểm tra thích ứng
Kết quả bài kiểm tra cố định	Pearson Correlation	1	.626**
	Sig. (2-tailed)		.000
	N	205	205
Kết quả bài kiểm tra thích ứng	Pearson Correlation	.626**	1
	Sig. (2-tailed)	.000	
	N	205	205
**. Correlation is significant at the 0.01 level (2-tailed).			

Như có thể thấy từ Bảng 3.19, giá trị Sig. (2-Tailed) là 0,000, nhỏ hơn 0,05. Có thể kết luận là mối tương quan có ý nghĩa thống kê giữa hai biến nghiên cứu. Hệ số 0,626 trong phân tích tương quan Pearson cho thấy mối tương quan dương, tương đối mạnh giữa điểm của bài kiểm tra cố định và điểm của bài kiểm tra thích ứng. Có thể hiểu rằng bài kiểm tra thích ứng cũng có thể phân loại mức độ kiến thức từ vựng tiếp nhận của thí sinh một cách đáng tin cậy, giống như bài kiểm tra cố định.

3.2.3. Phản hồi của thí sinh thực hiện trắc nghiệm thích ứng

3.2.3.1. Từ khảo sát

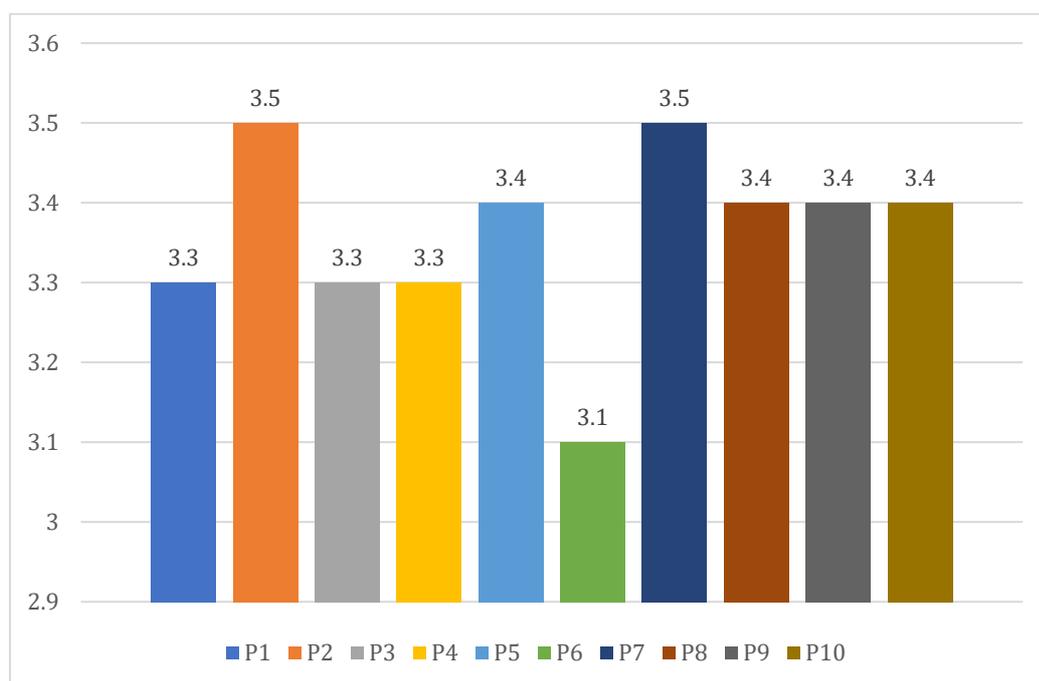
Trong phần này, các câu trả lời của người tham gia khảo sát sẽ được tổng hợp, phân tích và thảo luận để thể hiện nhận thức của thí sinh về bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh sau khi họ có những trải

nghiệm trực tiếp với bài kiểm tra thích ứng đánh giá kiến thức từ vựng tiếp nhận tiếng Anh.

Bảng 3.19: Nhận thức của học sinh về đặc điểm bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh

<i>Nội dung khảo sát</i>		4 Hoàn toàn đồng ý	3 Đồng ý một phần	2 Phản đối một phần	1 Hoàn toàn phản đối	Giá trị trội	Giá trị trung bình
<i>P1</i>	<i>Hiệu quả về thời gian</i>	28	40	5	1	3	3.3
<i>P2</i>	<i>Tính năng thân thiện với người dùng</i>	39	35	0	0	4	3.5
<i>P3</i>	<i>Đánh giá chính xác năng lực thí sinh</i>	25	46	3	0	3	3.3
<i>P4</i>	<i>Mức độ phù hợp của câu hỏi thi với năng lực của thí sinh</i>	27	44	3	0	3	3.3
<i>P5</i>	<i>Giảm lo lắng thi cử</i>	35	36	3	0	3	3.4
<i>P6</i>	<i>Giảm khả năng gian lận liên quan đến việc ghi nhớ bài, trao đổi đáp án</i>	23	38	11	2	3	3.1
<i>P7</i>	<i>Thông báo kết quả ngay lập tức</i>	40	34	0	0	4	3.5
<i>P8</i>	<i>Gợi ý học tập cho thí sinh ngay sau khi thi</i>	31	40	2	1	3	3.4
<i>P9</i>	<i>Khả năng sử dụng như một công cụ tự đánh giá hiệu quả</i>	30	40	4	0	3	3.4
<i>P10</i>	<i>Khả năng sử dụng như một công cụ hữu hiệu hỗ trợ việc học và ôn tập thường xuyên</i>	34	39	1	0	3	3.4

Như có thể thấy trong Bảng 3.20, những người tham gia hầu hết bày tỏ quan điểm tích cực của họ đối với bài thi trắc nghiệm thích ứng. Tất cả các tính năng của bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh đều nhận được mức đồng ý trung bình cao, cao hơn 3. Các tính năng thân thiện với người dùng và thông báo ngay lập tức về kết quả làm bài có phản hồi tốt nhất với điểm trung bình cao nhất – 3,5, đồng thời tiềm năng giảm khả năng gian lận chỉ đạt được giá trị trung bình thấp nhất – 3,1.



Hình 3.21: Giá trị trung bình mức độ đồng ý với các nhận định về bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh

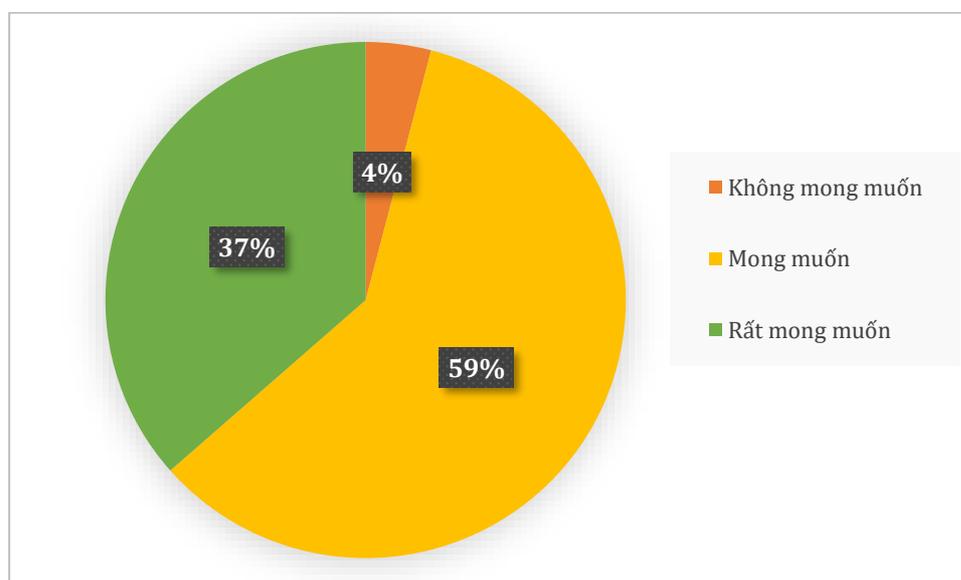
Với câu hỏi mở về điều gì khiến thí sinh hài lòng nhất khi trải nghiệm bài kiểm tra, phản hồi nhận được khá đa dạng. Mười một người tham gia khảo sát cho biết họ hài lòng với tất cả các khía cạnh của bài kiểm tra. Hai mươi người cho rằng đây là một công cụ hữu ích để học và ôn tập, đặc biệt là giúp nâng cao kiến thức từ vựng tiếng Anh của họ. Mười bốn người tham gia hài lòng nhất với đặc điểm thân thiện với người dùng của bài kiểm tra. Sáu người cho rằng đây là một công cụ tuyệt vời để đánh giá vốn từ vựng của họ, trong

đó có hai người nhấn mạnh tính chính xác trong việc xác định năng lực của thí sinh. Chất lượng của các câu hỏi cũng mang lại những trải nghiệm tích cực cho thí sinh; bảy người trả lời đánh giá cao câu hỏi dễ hiểu; hai người khen câu hỏi phù hợp với khả năng của ứng viên; 15 người cho rằng danh sách từ vựng được sử dụng trong bài thi rất thú vị và có giá trị. Ngoài ra, hiệu quả về thời gian và phản hồi nhanh chóng được đề cập là điểm mạnh của bài kiểm tra.

Sau đó, những người tham gia được hỏi về các yếu tố nào của bài kiểm tra khiến họ cảm thấy không hài lòng. 78,4% người tham gia khẳng định họ hài lòng với bài kiểm tra. Mười sáu người, trong đó có mười nam và sáu nữ, đã kể lại những trải nghiệm khó chịu. Hai người cho rằng bài kiểm tra có giao diện khá đơn giản, chưa hấp dẫn. Một người nghĩ rằng bài kiểm tra nên dài hơn với nhiều câu hỏi hơn. Năm người trong số họ, với vốn từ vựng ước tính dưới 2750 từ, phàn nàn rằng một số câu hỏi trong bài kiểm tra họ làm quá khó. Điều này có thể lý giải đơn giản là danh sách từ vựng được sử dụng trong bài thi rộng hơn những gì các em đã học và biết trước đó. Một số sinh viên bày tỏ mong đợi nhận được phản hồi với tất cả các câu trả lời đúng và sai cho các câu hỏi họ gặp trong bài kiểm tra và về việc phát triển ứng dụng có thể sử dụng trên điện thoại thông minh.

Câu hỏi cuối cùng là về mức độ mong muốn của thí sinh đối với các bài kiểm tra thích ứng trong tương lai. Các câu trả lời không có gì đáng ngạc nhiên. Với những phản hồi tích cực nêu trên, 71/74 người tham gia bày tỏ mong muốn hoặc rất mong muốn được tham gia các thử nghiệm thích ứng trong tương lai. Ba người không muốn tham gia các bài kiểm tra liên quan đến thuộc nhóm sử dụng điện thoại thông minh để làm bài kiểm tra. Phù hợp với những phản hồi trên, phát hiện này gợi ý rằng nhóm phát triển hệ thống kiểm tra thích ứng nên phát triển các ứng dụng có giao diện tương tác và hấp

dẫn hơn trên điện thoại thông minh để mang lại trải nghiệm tốt nhất cho người dự thi bất kể họ sử dụng thiết bị nào.



Hình 3.22: Mức độ mong muốn của người tham gia khảo sát về trải nghiệm với trắc nghiệm thích ứng trong tương lai

3.2.3.2. Từ phỏng vấn

Trong phần này, các câu trả lời từ các cuộc phỏng vấn được tóm tắt, phân tích và diễn giải để tìm hiểu thêm về quan điểm của thí sinh về bài kiểm tra thích ứng từ vựng tiếng Anh sau khi thí sinh trực tiếp làm bài. Tên của các sinh viên tham gia đã được thay đổi trong quá trình phân tích và diễn giải để đảm bảo tính bảo mật như đã cam kết trước đó với các sinh viên tham gia phỏng vấn.

Tất cả các cuộc phỏng vấn đều bắt đầu bằng phần giới thiệu của người được phỏng vấn để thu thập thông tin cơ bản của người tham gia. Các thí sinh đều là sinh viên năm nhất của một trường đại học; có ba nam và bảy nữ. Hầu hết các em tỏ ra không tự tin khi được yêu cầu tự đánh giá vốn từ vựng của mình và giải thích vì chưa làm bài kiểm tra từ vựng hoặc không sử dụng

một công cụ đánh giá cụ thể nào. Tất cả những người tham gia phỏng vấn đều thừa nhận rằng bài kiểm tra thích ứng từ vựng tiếng Anh là trải nghiệm đầu tiên với thử nghiệm thích ứng. Máy tính xách tay là thiết bị được họ sử dụng phổ biến nhất trong trải nghiệm (8/10 người tham gia), tiếp theo là điện thoại thông minh.

Cũng như ghi nhận từ kết quả khảo sát, những người tham gia đều bày tỏ quan điểm tích cực đối với bài kiểm tra thích ứng từ vựng tiếp nhận mà họ đã trải nghiệm, đồng thời đưa ra những nhận định cụ thể hơn về các tính năng của bài kiểm tra theo trải nghiệm của họ.

Độ chính xác của bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh

Khi được yêu cầu đánh giá độ chính xác của bài kiểm tra, một số thí sinh tỏ ra nghi ngờ rằng hình thức câu hỏi trong bài kiểm tra cho phép đoán mò, dù không biết câu trả lời vẫn có 25% khả năng đưa ra đáp án đúng, có thể ảnh hưởng đến độ chính xác của kết quả. Tuy nhiên, họ đều thừa nhận họ hài lòng với số điểm mình đạt được và tin rằng số điểm đó phản ánh đúng năng lực của mình. Một thí sinh tên NHẬT cho biết “Em đã làm ba bài khác nhau với các câu hỏi khác nhau; kết quả em nhận được gần như nhau nên em tin rằng hệ thống đã đánh giá chính xác năng lực của em”. Tương tự, hai thí sinh khác đã làm nhiều lượt trên hệ thống, NGA và CẨM đã thể hiện sự tin tưởng vào tính chính xác của bài kiểm tra thích ứng trong việc đánh giá từ vựng dựa trên số điểm họ đạt được.

Hiệu quả thời gian của bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh

Các thí sinh đều phàn nàn về độ dài của bài thi cố định 100 câu và đồng ý rằng thời gian làm bài kiểm tra thích ứng giảm đáng kể và hợp lý hơn so với bài thi cố định. Việc giảm thời gian này mang lại cho họ những trải

nghiệm tốt hơn trong việc đánh giá từ vựng. Hai người trong số họ, NGA và LAN mong đợi nhiều câu hỏi hơn trong bài kiểm tra thích ứng; họ gợi ý 40-50 câu hỏi trong mỗi bài kiểm tra. Ngoài ra, THANH nhấn mạnh thời gian làm bài thi một phần phụ thuộc vào chất lượng đường truyền internet. Nhìn chung, những người tham gia phỏng vấn đánh giá cao hiệu quả về mặt thời gian của bài kiểm tra thích ứng.

Mức độ thân thiện với người dùng của bài kiểm tra thích ứng

Mặc dù có sự khác biệt về thiết bị được sử dụng trong trải nghiệm với bài kiểm tra thích ứng, các thí sinh tham gia phỏng vấn vẫn tỏ ra hài lòng nhờ thao tác làm bài đơn giản. Việc lựa chọn các câu hỏi của hệ thống phù hợp với năng lực của thí sinh được đánh giá cao. NGA so sánh trải nghiệm của mình trong bài kiểm tra cố định và trắc nghiệm thích ứng “Khi làm đề 100 câu, đoạn đầu em làm nhanh lắm, nhưng đoạn sau phải dừng lại nghĩ. Đến trắc nghiệm thích ứng thì thấy khá là ổn, khi em thấy làm câu khó mà có đáp án sai thì câu hỏi sau lại thấy dễ hơn, khá phù hợp với năng lực của mình”. Cùng quan điểm, NHI cho biết “Em không thấy có kiểu một loạt các câu hỏi quá khó hoặc quá dễ khiến em không muốn làm, các câu hỏi có vẻ gần và hợp với năng lực của em”.

Tất cả các thí sinh đều đề cập đến trải nghiệm làm bài dễ chịu, cảm giác thoải mái khi làm bài kiểm tra thích ứng và đưa ra lý do riêng, bao gồm việc lựa chọn câu hỏi phù hợp năng lực thí sinh, thời gian làm bài linh hoạt và không bị áp lực như khi làm bài kiểm tra trên lớp.

Ngoài ra, tất cả các thí sinh tham gia phỏng vấn đưa ra những nhận định rất tích cực với việc được nhận kết quả và gợi ý ôn tập ngay sau khi thực hiện bài kiểm tra. THU khẳng định “Em thấy thế là rất ổn, ngay sau khi làm mình biết mình ở mức nào, có thể cải thiện được không. Theo em thì ai cũng

thích biết kết quả ngay sau khi làm bài”. Đồng quan điểm, CÀM cho biết “Em nghĩ rất là hữu ích đó ạ, giúp em học bài đúng hướng hơn, nếu như những bạn có nhu cầu đặc biệt thì lại càng hiệu quả hơn, như kiểu bỏ khuyết, bỏ sung những cái mình thiếu”. Có thể thấy, việc có được kết quả làm bài ngay sau khi thí sinh hoàn thành đã mang lại những phản hồi rất tốt từ thí sinh, góp phần không nhỏ vào tính hiệu quả của hệ thống nói chung, cũng như bài kiểm tra từ vựng nói riêng.

Tính hữu dụng của bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh

Những thí sinh tham gia phỏng vấn từ những trải nghiệm trực tiếp của mình với bài kiểm tra thích ứng tin tưởng rằng trắc nghiệm thích ứng có thể giảm các tiềm năng gian lận thi cử liên quan đến việc nhớ đề hay nhắc bài giữa các thí sinh. Tất cả các thí sinh đều có cùng quan điểm như thí sinh tên NHẬT “thí sinh khác nhau sẽ làm các bài kiểm tra thích ứng trên máy tính khác nhau. Ngay với mỗi thí sinh thì các lượt làm bài cũng khác nhau nên chắc sẽ hạn chế một phần gian lận, vì số lượng câu lặp vẫn có nhưng không nhiều nên không thể nhắc nhau hay nhớ đề”. Những ý kiến này cùng với các mô tả về quá trình làm bài của thí sinh trên hệ thống được trình bày ở trên khẳng định tính ưu việt của trắc nghiệm thích ứng. Với ngân hàng câu hỏi đủ lớn và các thuật toán thích ứng, trắc nghiệm thích ứng đã mang lại những bài kiểm tra và các câu hỏi đa dạng, giảm thiểu sự trùng lặp từ đó tăng thêm hiệu quả trong kiểm tra đánh giá.

Cũng nhờ tính ưu việt của trắc nghiệm thích ứng, các thí sinh tham gia phỏng vấn cho biết việc sử dụng trắc nghiệm thích ứng như một công cụ tự đánh giá cũng như một công cụ để học tập và ôn luyện. NGÀ coi bài kiểm tra thích ứng như một công cụ hỗ trợ hữu ích để tự đánh giá năng lực của bản thân bên cạnh các hình thức đánh giá trên lớp, “Theo hướng tích cực, việc tự

đánh giá sẽ giúp mình có phương án để tự biết mình thế nào, tự điều chỉnh, tự ôn luyện, không cần phụ thuộc vào giáo viên. Nhiều khi trên lớp giáo viên mà đông học sinh ấy thì không thể sát sao với từng cá nhân và đánh giá chuẩn xác được”. Trong khi đó, THANH có gợi ý cách thức sử dụng trắc nghiệm thích ứng trong quá trình tự học để nâng cao kiến thức và năng lực, “Về mảng từ vựng, em nghĩ nó là một công cụ học tập rất hữu ích, ví dụ có thể phát triển theo cách là theo dõi quá trình học được những từ nào rồi, sau đó tạo lập một sơ đồ cây, cho lên đó những từ vựng em sẽ học, liên quan đến và có thể đạt được, kiểu như thế khi em làm xong trắc nghiệm em có thể học những từ liên quan và giúp ích nhiều hơn”. Có thể thấy qua các câu trả lời phỏng vấn là quan điểm tích cực và lạc quan của thí sinh với việc áp dụng trắc nghiệm thích ứng trong quá trình học tập và rèn luyện của bản thân, trắc nghiệm thích ứng từ đó có giá trị hơn khi tạo động lực cho quá trình tự học, tự rèn luyện của người học.

Một số đề xuất để tăng hiệu quả của bài kiểm tra

Tất cả thí sinh tham gia phỏng vấn có quan điểm đồng nhất về tính hiệu quả của bài kiểm tra thích ứng từ vựng mà họ có trải nghiệm trực tiếp. Một số thể hiện rõ sự hài lòng với các tính năng vượt trội của trắc nghiệm thích ứng so với bài kiểm tra cố định truyền thống. Không khác biệt nhiều với kết quả từ khảo sát, những tính năng được khen ngợi nhiều nhất, mang lại sự hài lòng cho thí sinh trong các câu trả lời phỏng vấn là thao tác thực hiện đơn giản và kết quả tức thì. Hai tính năng này được xem là rất thân thiện với thí sinh, hứa hẹn có thể hướng tới các đối tượng khác nhau, không đòi hỏi kỹ năng đặc biệt gì và việc nhận kết quả tức thì vừa có giá trị tâm lý vừa giúp thí sinh xác định năng lực cũng như hướng học tập phù hợp. Ngoài ra, thời gian ngắn và danh sách từ vựng được kiểm tra cũng được nhiều thí sinh cảm nhận tốt và đánh giá cao.

Với mong muốn được sử dụng nhiều hơn trắc nghiệm thích ứng trong tương lai, các thí sinh tham gia phỏng vấn đưa ra những đề xuất về giao diện, về nội dung tùy thuộc vào nhu cầu của bản thân. Các đề xuất khá đa dạng và nhóm phát triển hệ thống có thể lưu ý để nâng cấp các tính năng của hệ thống và mở rộng nội dung. Các đề xuất phần nhiều hướng tới việc phát triển các nội dung mới liên quan đến việc học tiếng Anh. Với từ vựng, các thí sinh NHẬT, NHI, NGA, THU và CÂM gợi ý đến từ vựng chuyên ngành, từ vựng theo chủ đề hay từ vựng phục vụ các kỳ thi quốc tế như TOEIC hay IELTS. Riêng THU có thêm gợi ý về việc khai thác các khía cạnh khác của từ bên cạnh ngữ nghĩa để hướng tới việc sử dụng từ. Các thí sinh TÂM, THU và TÚ mong đợi trắc nghiệm thích ứng sẽ hướng tới kỹ năng nghe hay các nội dung giao tiếp – những e ngại của hai em trong quá trình học tiếng Anh, trong khi THANH thì cho rằng trắc nghiệm thích ứng phù hợp nhất với từ vựng và ngữ pháp, nên hướng phát triển khả thi nhất là ngữ pháp, cụ thể là thì trong tiếng Anh. Về tính năng của hệ thống, THU trình bày mong muốn tăng sự hấp dẫn cho hệ thống với các âm thanh và hình ảnh “ngoài việc chèn hình ảnh có thể thêm video hay file ghi âm, việc có thêm hình ảnh âm thanh thú vị sẽ không chỉ thu hút sinh viên mà còn cả các em cấp 1 và cấp 2, thu hút nhiều đối tượng hơn”. NGA và LAN muốn được xem luôn đáp án sau mỗi câu hỏi để có thể học luôn từ mới. BÁCH muốn hệ thống “cho thêm các lựa chọn về số lượng câu hỏi như bài kiểm tra 20 câu, 30 câu, 50 câu hay 100 câu tùy theo thời gian mà người dùng có”. Thú vị nhất có thể là VÂN, chuyên ngành Toán Tin, trình bày sự hứng thú với trắc nghiệm thích ứng và tính ứng dụng của trắc nghiệm thích ứng, mong được hiểu biết nhiều hơn nữa về trắc nghiệm thích ứng vì “có thể sau này em cũng làm những cái như thế này”.

Các phản hồi về trắc nghiệm thích ứng trong nghiên cứu có nhiều điểm chung với những phát hiện của Apostolou và cộng sự (2009) và Hussain và

cộng sự (2021). Tuy nhiên, chúng rất khác với quan điểm của Kimura (2017) về việc giảm động lực và hiệu quả học tập. Có thể giải thích rằng việc sử dụng CAT với những mục đích khác nhau trong những bối cảnh khác nhau sẽ dẫn đến những góc nhìn khác nhau từ các ứng viên. Tuy nhiên, không thể phủ nhận những ưu điểm vượt trội của CAT hay các bài thi trên máy tính như thao tác làm bài dễ dàng, phản hồi nhanh chóng sau khi làm bài luôn được các thí sinh quan tâm. Hơn nữa, những nghiên cứu về nhận thức của người làm bài thi này có giá trị rất lớn trong quá trình phát triển bài thi nhằm mang lại kết quả đánh giá thân thiện và thiết thực hơn với người dùng. Vì hầu hết thí sinh nhận thấy bài kiểm tra có liên quan và phù hợp với kiến thức từ vựng tiếng Anh của họ nên nghiên cứu này cung cấp bằng chứng để tăng độ tin cậy của bài kiểm tra như một công cụ đánh giá ngôn ngữ có giá trị và đáng tin cậy. Khám phá nhận thức của thí sinh cũng giúp phát hiện ra những sai lệch tiềm ẩn trong việc đánh giá cũng như những cải tiến cần thiết của hệ thống CAT. Cuối cùng nhưng không kém phần quan trọng, việc điều tra nhận thức của thí sinh có thể góp phần vào lĩnh vực nghiên cứu về trải nghiệm thích ứng trên máy tính và đánh giá ngôn ngữ.

Kết quả rất khả quan, góp phần vào lĩnh vực nghiên cứu về phát triển và đánh giá các bài kiểm tra thích ứng trên máy tính ở Việt Nam. Những đề xuất của những người tham gia khảo sát và phỏng vấn về việc nâng cấp hệ thống cũng cần được xem xét trong quá trình phát triển, bảo trì và cải tiến các bài kiểm tra thích ứng trong tương lai để có thêm những phản hồi tích cực từ thí sinh. Ngoài ra, nhóm phát triển khóa học và các bên liên quan nên cân nhắc tiến hành thêm nhiều nghiên cứu đánh giá nhận thức của người dự thi, từ đó đưa ra nhận định tổng thể về hệ thống kiểm tra thích ứng đã phát triển.

3.3. Bàn luận

Phần 3.1 báo cáo kết quả xây dựng ngân hàng câu hỏi chuẩn hóa để kiểm tra từ vựng tiếp nhận tiếng Anh cho đối tượng người học tiếng Anh ở Việt Nam. Sau khi xem xét kết quả phân tích của bảy đề, nhóm tác giả đi đến kết luận rằng 30 câu hỏi thô không phù hợp với mô hình đang sử dụng nên sẽ bị loại bỏ, không đưa vào ngân hàng câu hỏi. Bên cạnh đó, 84 câu hỏi yêu cầu chỉnh sửa các phương án nhiễu để nâng cao chất lượng câu hỏi. Ngân hàng sau khi cân bằng có tổng 522 câu hỏi, đã được định cỡ trên cùng một thang đo để có thể sử dụng trong hệ thống trắc nghiệm thích ứng UEd-CAT.

Phần kết quả này, cùng với các nghiên cứu trước đây về phát triển và xác trị đề kiểm tra (Le và cộng sự, 2019; Lê Thái Hưng và cộng sự, 2019; Lê Thái Hưng & Nguyễn Thái Hà, 2021; Nguyen & Nguyen, 2020; Nguyễn Thái Hà và cộng sự, 2021), mang lại những đóng góp tích cực trong lĩnh vực kiểm tra đánh giá tại Việt Nam, nhấn mạnh tầm quan trọng của một quy trình xây dựng đề thi nghiêm túc để đảm bảo độ giá trị và độ tin cậy của các công cụ kiểm tra đánh giá cũng như khuyến khích sử dụng các công cụ phân tích để hỗ trợ các nhà giáo dục cũng như các nhà nghiên cứu thực hiện các nghiên cứu xác trị một cách hiệu quả. Những kết quả tích cực này có thể tạo tiền đề cho việc thiết kế các ngân hàng câu hỏi có giá trị và đáng tin cậy để phục vụ các mục đích và nhu cầu khác nhau của người học, giáo viên và nhà nghiên cứu trong bối cảnh giáo dục tại Việt Nam.

Phần 3.2 trình bày kết quả phục vụ mục đích khảo sát tính khả thi của việc tạo ra một công cụ đánh giá từ vựng tiếp nhận tiếng Anh trong hệ thống UEd-CAT hướng tới người học ngoại ngữ tiếng Anh ở Việt Nam. Sử dụng các thuật toán thích ứng có sẵn, nhà nghiên cứu đã áp dụng các đặc tả của NGSLT và phát triển các câu hỏi song ngữ Anh – Việt để xây dựng một ngân hàng câu

hỏi chuẩn hóa lớn và tiến hành thử nghiệm với đối tượng sinh viên tại một trường đại học ở Việt Nam để phục vụ mục đích nghiên cứu.

Sau khi thử nghiệm, khảo sát và phỏng vấn, quá trình và kết quả làm bài của những người tham gia cũng như quan điểm của họ đã được ghi lại và phân tích để đánh giá tính hợp lệ và độ tin cậy của công cụ đánh giá trong việc đo lường kiến thức từ vựng tiếp nhận tiếng Anh so với bài kiểm tra cố định.

Các phát hiện cho thấy bài kiểm tra thích ứng từ vựng sở hữu những đặc điểm nổi bật đáng chú ý. Trước hết, với cùng mục tiêu như đề cố định NGSLT, bài kiểm tra thích ứng từ vựng có thể xác định kiến thức từ vựng tiếp nhận của thí sinh với ít câu hỏi hơn, từ 6 đến 20 câu hỏi thay vì một trăm câu hỏi, chỉ cần khoảng 1/5 số câu hỏi kiểm tra so với bài kiểm tra cố định. Do đó, cả thời gian và công sức của thí sinh đều giảm đi. Hiệu quả thời gian này của bài kiểm tra thích ứng từ vựng cung cấp thêm bằng chứng cho thấy hiệu quả tiết kiệm thời gian là một thế mạnh của CAT trong thực tiễn đánh giá giáo dục cùng với các nghiên cứu trước đây (Giourogrou & Economides, 2003; Khoshsima & Toroujeni, 2017; Meunier, 1994; Mizumoto và cộng sự, 2019; Okhotnikova và cộng sự, 2019; Pathan, 2012; Tseng, 2016).

Thứ hai, các thuật toán thích ứng có thể giúp chọn các câu hỏi phù hợp với năng lực của thí sinh từ ngân hàng câu hỏi và cung cấp các câu hỏi cũng như gói câu hỏi khác nhau với tỷ lệ trùng lặp thấp giữa các lần làm bài khác nhau của một thí sinh. Tính năng này có thể nâng cao trải nghiệm của thí sinh vì nó có thể giúp thí sinh giảm bớt căng thẳng và buồn chán trong quá trình làm bài (Giourogrou & Economides, 2003; Rasskazova và cộng sự, 2017).

Hơn nữa, tương tự như phiên bản NGSLT cố định, bài kiểm tra thích ứng cung cấp các điểm số có thể được diễn giải để xác định trình độ từ vựng tiếp nhận của thí sinh, một khía cạnh cơ bản đánh giá mức độ thành thạo ngôn ngữ.

Phân tích tương quan các kết quả trong bài kiểm tra từ vựng cố định và thích ứng cho thấy độ tin cậy cao của trắc nghiệm thích ứng so với trắc nghiệm cố định. Nói cách khác, bài kiểm tra từ vựng tiếp nhận thích ứng có thể được sử dụng như một giải pháp thay thế đáng tin cậy cho bài kiểm tra cố định 100 câu với độ chính xác cao, giúp định hướng người học ngoại ngữ tiếng Anh ở Việt Nam nắm vững danh sách từ vựng NSGL gồm 2801 từ tiếng Anh thông dụng, từ đó tạo nền tảng vững chắc cho việc dạy và học ngoại ngữ hiệu quả (Webb & Chang, 2012). Điều này cũng nhất quán với kết quả của các nghiên cứu trước đây về độ chính xác của trắc nghiệm thích ứng trong đánh giá ngôn ngữ (Choi và cộng sự, 2003; Gawliczek và cộng sự, 2021; Giourogrou & Economides, 2003; He & Min, 2017; Khoshsima & Toroujeni, 2017; Larson & Madsen, 1985; Meunier, 1994; Mizumoto và cộng sự, 2019; Pathan, 2012).

Tính hữu ích của bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh cũng được khám phá trong nghiên cứu này. Cùng với các bài kiểm tra thích ứng khác như CARPRO (Tseng, 2016), CATSS (Aviad-Levitzky và cộng sự, 2019), và CAT-WPLT (Mizumoto và cộng sự, 2019), bài kiểm tra thích ứng từ vựng được xây dựng trong luận án chứng minh rằng CAT cung cấp các tính năng ưu việt và hiệu quả để đánh giá năng lực từ vựng tiếp nhận tiếng Anh. Hướng về tương lai, nhờ những phát hiện tích cực từ các nghiên cứu về CALT, việc phát triển các công cụ đánh giá ngôn ngữ thích ứng có thể được đa dạng hóa để bao gồm nhiều lĩnh vực ngôn ngữ khác nhau. Ban đầu, các bài đánh giá từ vựng theo chủ đề có thể được giới thiệu, phục vụ cho các chủ đề cụ thể hoặc các mục đích thực tế khác nhau. Sau đó, các đánh giá có thể liên quan đến việc đánh giá ngữ pháp và cách phát âm của người học ngoại ngữ tiếng Anh, cũng như các kỹ năng tiếp nhận như đọc và nghe. Mục tiêu cuối cùng là cung cấp cho người học quyền truy cập miễn phí vào các công cụ hiệu quả và đóng góp

vào lĩnh vực đánh giá ngôn ngữ nhằm phát huy tiềm năng của CALT trong bối cảnh dạy và học ngoại ngữ tiếng Anh.

Tương tự như các công cụ đánh giá khác trong hệ thống UEd-CAT, bài kiểm tra thích ứng từ vựng có thể đóng vai trò là phương tiện để tự đánh giá và thực hành thường xuyên (Le và cộng sự, 2019; Le & Nguyen, 2021; Nguyen & Le, 2018). Từ phản hồi của những người tham gia khảo sát và phỏng vấn có thể thấy rằng bài kiểm tra thích ứng từ vựng nhận được những tín hiệu rất tích cực từ phía thí sinh – những người trực tiếp trải nghiệm bài kiểm tra. Từ góc nhìn của nhiều thí sinh, bài thi có thể là một công cụ đánh giá và học tập hữu ích giúp họ nâng cao kiến thức và năng lực, đặc biệt là từ vựng tiếng Anh. Bên cạnh những điểm mạnh vượt trội về nội dung kiểm tra, tính năng kiểm tra, chất lượng bài thi, hiệu quả về thời gian và phản hồi theo thời gian thực, bài kiểm tra vẫn phải đối mặt với thách thức về việc cải thiện giao diện, nội dung phản hồi hấp dẫn và cần có phiên bản dành cho điện thoại thông minh. Bài kiểm tra thích ứng từ vựng cũng được kỳ vọng sẽ góp phần vào việc học tập thích ứng của người học để có hiệu suất và kết quả tốt hơn. Ngoài ra, việc cung cấp các lựa chọn để học sinh kiểm tra và học từ vựng một cách độc lập có thể nuôi dưỡng các kỹ năng học tập tự định hướng thiết yếu và quý giá cho việc học tập suốt đời.

Các công cụ đánh giá tích hợp trải nghiệm thích ứng bằng máy tính cũng được tin tưởng rằng không chỉ trao quyền cho người học kiểm soát quá trình học tập của mình mà còn cung cấp cho các nhà giáo dục những hiểu biết sâu sắc có giá trị để có chiến lược giảng dạy hiệu quả hơn. Các nhà nghiên cứu và nhà phát triển hệ thống cũng luôn ý thức rằng CAT yêu cầu những nỗ lực xác trị, duy trì và đổi mới liên tục để phục vụ nhu cầu luôn thay đổi của thực tiễn giáo dục để CAT có thể duy trì hiệu quả và dễ tiếp cận cho tất cả mọi người, bất kể hoàn cảnh hoặc thách thức họ đang đối mặt.

3.4. Kết chương

Chương 3 đã trình bày hai phần kết quả nghiên cứu bao gồm ngân hàng câu hỏi trắc nghiệm thích ứng kiểm tra từ vựng tiếp nhận tiếng Anh đã được chuẩn hóa đáp ứng yêu cầu kỹ thuật để được sử dụng trong hệ thống trắc nghiệm thích ứng UEd-CAT; và kết quả thử nghiệm của thí sinh với bài trắc nghiệm thích ứng từ vựng tiếp nhận tiếng Anh trên hệ thống Ued-CAT của trường ĐHGĐ-ĐHQGHN. Các kết quả thu được đã được tổng hợp, phân tích và bàn luận với sự kết nối và so sánh với các nghiên cứu liên quan trước đây cả trong nước và ngoài nước để đi đến những đánh giá về bài trắc nghiệm thích ứng từ vựng tiếp nhận tiếng Anh dành cho đối tượng người học ngoại ngữ tiếng Anh ở Việt Nam, từ đó hướng tới việc khẳng định những đóng góp của luận án, những hướng đi cho các nghiên cứu tiếp theo cũng như khuyến nghị tới các bên liên quan sẽ được trình bày trong phần kết luận.

KẾT LUẬN

1. Tóm lược kết quả nghiên cứu của luận án

Luận án được thực hiện với mục đích phát triển bài kiểm tra thích ứng bằng máy tính để đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của người học ngoại ngữ tiếng Anh tại Việt Nam, cụ thể luận án đã thực hiện các nhiệm vụ nghiên cứu và trả lời hai câu hỏi nghiên cứu:

Câu hỏi 1: Ngân hàng câu hỏi trắc nghiệm đánh giá kiến thức từ vựng tiếp nhận tiếng Anh được xây dựng và chuẩn hóa như thế nào?

Câu hỏi 2: Bài kiểm tra thích ứng bằng máy tính được thiết kế thực hiện việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của người học ngoại ngữ tiếng Anh ở Việt Nam như thế nào?

Kết quả nghiên cứu thu nhận được đã hoàn thành việc trả lời hai câu hỏi nghiên cứu. Từ việc tổng quan cơ sở lý thuyết và các nghiên cứu đã được thực hiện về trắc nghiệm thích ứng và kiểm tra từ vựng tiếp nhận tiếng Anh, luận án có thể đi đến kết luận trắc nghiệm thích ứng đã và đang được áp dụng trong kiểm tra đánh giá ngôn ngữ, cụ thể là đánh giá kiến thức từ vựng tiếng Anh, là một xu hướng phát triển áp dụng công nghệ máy tính trong giáo dục được đón nhận trên thế giới cũng như tại Việt Nam. Bài kiểm tra thích ứng từ vựng được thiết kế trong luận án là phù hợp với xu hướng. Việc sử dụng lý thuyết hỏi đáp trong thiết kế công cụ đánh giá đơn năng lực, cụ thể trong luận án là kiến thức từ vựng tiếp nhận tiếng Anh, là một nỗ lực đáng ghi nhận trong bối cảnh dạy và học ngoại ngữ tiếng Anh ở Việt Nam cũng như trong bối cảnh giáo dục tại Việt Nam nói chung.

Với câu hỏi nghiên cứu thứ nhất, luận án đã hoàn thành mục tiêu xây dựng ngân hàng câu hỏi chuẩn hóa để kiểm tra từ vựng tiếp nhận tiếng Anh cho đối tượng người học tiếng Anh ở Việt Nam. Việc xây dựng ngân hàng câu hỏi

được tiến hành công phu với nhiều giai đoạn, từ việc xây dựng đề mẫu, viết câu hỏi thô, chỉnh sửa theo phương pháp chuyên gia, thử nghiệm với thí sinh và phân tích kết quả với mô hình IRT một tham số. Sau khi xem xét kết quả phân tích, luận án đã thu nhận kết quả rằng 30 câu hỏi thô không phù hợp với mô hình đang sử dụng nên sẽ bị loại bỏ, không đưa vào ngân hàng câu hỏi, và 84 câu hỏi yêu cầu chỉnh sửa các phương án nhiễu để nâng cao chất lượng câu hỏi. Ngân hàng câu hỏi, sau khi áp dụng phương pháp cân bằng, có tổng 522 câu hỏi với các tham số đã được định cỡ trên cùng một thang đo để có thể sử dụng trong hệ thống trắc nghiệm thích ứng UEd-CAT.

Với câu hỏi nghiên cứu thứ hai, luận án đã tiến hành thử nghiệm bài kiểm tra thích ứng từ vựng tiếp nhận tiếng Anh với đối tượng thí sinh hướng tới là người học ngoại ngữ tiếng Anh ở Việt Nam, cụ thể là đối tượng sinh viên đại học có bậc 1-2-3 trong Khung năng lực ngoại ngữ Việt Nam. Quá trình làm bài, kết quả làm bài cũng như quan điểm của thí sinh đã được tổng hợp và phân tích để đánh giá tính hiệu quả của bài kiểm tra luận án thiết kế. Các kết quả chỉ ra rằng bài kiểm tra có hiệu quả trong việc đánh giá kiến thức từ vựng tiếp nhận của thí sinh trong khi thời gian làm bài ngắn hơn, các câu hỏi đa dạng và phù hợp với năng lực thí sinh, bài trắc nghiệm nhìn chung mang lại trải nghiệm tích cực đến cho thí sinh và nhận lại những phản hồi khả quan từ các đối tượng tham gia nghiên cứu.

Tóm lại, với việc áp dụng các phương pháp nghiên cứu khoa học, luận án đã được tiến hành cẩn trọng và ghi nhận được những kết quả có giá trị, hoàn thành mục tiêu nghiên cứu đã đề ra.

2. Đóng góp của luận án

Về mặt lý luận, luận án là một công trình nghiên cứu khoa học được thực hiện dựa trên cơ sở lý thuyết chặt chẽ và tổng quan nghiên cứu có hệ thống khoa học về kiểm tra từ vựng và trắc nghiệm thích ứng bằng máy tính.

- Từ việc tổng quan các nghiên cứu và các bài kiểm tra từ vựng tiếng Anh đang được sử dụng, luận án có thể đi đến kết luận các bài kiểm tra từ vựng đều có giá trị thực tiễn và đang được sử dụng rộng rãi trong dạy và học tiếng Anh cũng như trong nghiên cứu. Việc lựa chọn một bài kiểm tra tùy thuộc vào mục đích sử dụng khác nhau của người học, người dạy và các nhà nghiên cứu, cũng như mức độ quen thuộc của người học với các dạng thức kiểm tra, từ đó mang lại những trải nghiệm tích cực cũng như kết quả thể hiện chính xác năng lực của thí sinh. Ngoài ra, qua việc tổng lược các bài kiểm tra từ vựng tiếp nhận tiếng Anh, luận án cũng nhấn mạnh tầm quan trọng của việc phát triển các công cụ kiểm tra đánh giá mới phục vụ các nhu cầu không ngừng thay đổi của cả người dạy và người học cũng như các nhà nghiên cứu, đồng thời đề xuất một số vấn đề cần được quan tâm trong các dự án về kiểm tra từ vựng trong tương lai, cụ thể là các vấn đề về xác định mục đích kiểm tra rõ ràng, lấy mẫu từ kiểm tra phù hợp mục đích kiểm tra, xác định định dạng câu hỏi phù hợp, phát triển phiên bản song ngữ, áp dụng công nghệ máy tính trong thiết kế cũng như trong thực hiện kiểm tra, và tiến hành xác trị nghiêm túc.

- Với việc tổng quan các nghiên cứu trước đây về trắc nghiệm thích ứng trong đánh giá và đào tạo ngôn ngữ, luận án đã nỗ lực tổng hợp những tiềm năng và những thách thức trong việc triển khai CALT trong đánh giá ngôn ngữ. CALT là một cách tiếp cận trong lĩnh vực đo lường và đánh giá trong giáo dục dựa trên máy tính nhằm điều chỉnh các bài kiểm tra ngôn ngữ phù hợp với mức độ thành thạo của người làm bài kiểm tra. Bên cạnh nhiều lợi ích của hệ thống kiểm tra tích hợp sử dụng máy tính, sáng kiến CALT còn có những ưu điểm rõ ràng so với các phương thức kiểm tra thông thường, như độ tin cậy được cải thiện, tiềm năng tiết kiệm thời gian, thúc đẩy trải nghiệm kiểm tra của thí sinh và các tùy chọn kiểm tra đa dạng. Tuy nhiên, CALT phải đối mặt với một số thách thức liên quan đến quá trình thiết kế và quản lý bài kiểm tra. Nhìn chung,

có thể tin tưởng mạnh mẽ rằng CALT sẽ ngày càng trở nên phổ biến như một lựa chọn đánh giá hiệu quả trong cả nghiên cứu và thực tiễn ở Việt Nam.

Về mặt thực tiễn, luận án đã thành công trong việc xây dựng và phát triển công cụ đánh giá kiến thức từ vựng tiếp nhận của người học ngoại ngữ tiếng Anh ở Việt Nam.

- Luận án cũng đã hoàn thành việc phát triển và cung cấp những bằng chứng xác trị đầu tiên với phiên bản song ngữ Anh Việt của bài kiểm tra từ vựng tiếng Anh thông dụng NGS LT, một đóng góp đáng ghi nhận trong lĩnh vực kiểm tra từ vựng tiếng Anh trong bối cảnh dạy và học ngoại ngữ tại Việt Nam, và lĩnh vực kiểm tra ngôn ngữ nói chung.

- Nghiên cứu của luận án, cùng với các nghiên cứu trước đây về xây dựng ngân hàng câu hỏi đánh giá các nội dung khác nhau được sử dụng trong hệ thống UEd-CAT mang lại những đóng góp tích cực trong lĩnh vực kiểm tra đánh giá tại Việt Nam, nhấn mạnh tầm quan trọng của một quy trình xây dựng đề thi nghiêm túc để đảm bảo độ giá trị và độ tin cậy của các công cụ kiểm tra đánh giá cũng như khuyến khích sử dụng các công cụ phân tích để hỗ trợ các nhà giáo dục cũng như các nhà nghiên cứu thực hiện các nghiên cứu xác trị một cách hiệu quả. Các kết quả thu được cũng đóng góp bằng chứng xác thực về việc áp dụng lý thuyết hồi đáp và phương pháp cân bằng trong xây dựng và chuẩn hóa ngân hàng câu hỏi trong lĩnh vực đo lường và đánh giá trong giáo dục, tạo tiền đề cho việc thiết kế các ngân hàng câu hỏi có giá trị và đáng tin cậy để phục vụ các mục đích và nhu cầu khác nhau của người học, giáo viên và nhà nghiên cứu trong bối cảnh giáo dục tại Việt Nam.

- Luận án cũng đã thu nhận những kết quả khả quan từ thử nghiệm, khảo sát và phỏng vấn để chỉ ra rằng bài kiểm tra thích ứng được thiết kế có hiệu quả trong việc đánh giá kiến thức từ vựng tiếp nhận tiếng Anh của thí sinh trong khi sử dụng ít câu hỏi hơn so với các bài kiểm tra có dạng cố định. Bản chất

thích ứng của công cụ này cũng mang lại trải nghiệm cá nhân hóa tích cực cho người làm bài kiểm tra, khi thí sinh gặp những câu hỏi phù hợp với trình độ năng lực của mình. Từ đó, bài kiểm tra từ vựng đã thành công trong việc mở rộng nội dung của hệ thống UEd-CAT ngoài phạm vi hiện có. Thêm vào đó, kết quả tích cực của nghiên cứu này đã mở đường cho việc ứng dụng CAT rộng rãi hơn trong thực tiễn giáo dục với các nội dung và mục đích khác nhau.

3. Khuyến nghị

Luận án đã khẳng định hiệu quả của trắc nghiệm thích ứng trong đo lường và đánh giá từ vựng cũng như trong quá trình dạy và học ngoại ngữ tiếng Anh, rộng hơn là việc áp dụng công nghệ trong nghiên cứu và thực tiễn lĩnh vực giáo dục. Từ những kết luận của luận án, tác giả xin đưa ra những khuyến nghị hướng tới các nhóm đối tượng liên quan như sau:

3.1. Khuyến nghị với người học

Trước hết là đối tượng sinh viên, đối tượng tham gia thử nghiệm của luận án: các em có thể tiếp tục sử dụng bài kiểm tra thích ứng như một công cụ tự đánh giá, cũng như công cụ ôn luyện để nắm vững danh sách từ vựng tiếng Anh thông dụng, sau đó là mở rộng kiến thức từ vựng với các khía cạnh khác, ngoài kiến thức từ vựng tiếp nhận, từ đó làm giàu kiến thức từ vựng cũng như năng lực ngôn ngữ của bản thân. Khi đã nhận biết những ưu điểm của công cụ trắc nghiệm thích ứng, các em cũng có thể giới thiệu đến các bạn học, hay các đối tượng khác có mong muốn tìm kiếm một công cụ để củng cố và mở rộng vốn từ vựng của mình.

Rộng hơn với đối tượng người học ngoại ngữ ở Việt Nam, người học cần có một thái độ nghiêm túc với việc học của mình, ý thức được tầm quan trọng của từ vựng, bền bỉ học tập và luyện tập, trước hết là tăng lượng từ vựng, kiến

thức từ vựng để từ đó phát triển các kỹ năng cũng như tăng cường mức độ thành thạo ngôn ngữ.

3.2. Khuyến nghị với giáo viên và các cơ sở đào tạo

Khi áp dụng công nghệ máy tính và cụ thể là trắc nghiệm thích ứng đã trở thành xu hướng phát triển trong giáo dục và đào tạo, giáo viên cần tự cập nhật những ưu điểm vượt trội của công nghệ và tích cực áp dụng trong thực tiễn giáo dục của mình. Giáo viên có thể giới thiệu trắc nghiệm thích ứng với người học, thực hiện các nghiên cứu đánh giá cũng như áp dụng trắc nghiệm thích ứng trong quá trình dạy của mình, đồng viên khuyến khích người học sử dụng để nâng cao kiến thức và rèn luyện kỹ năng.

Các cơ sở giáo dục cũng có thể đầu tư xây dựng hệ thống trắc nghiệm thích ứng phục vụ những nội dung và mục đích đào tạo của riêng mình, tích hợp trắc nghiệm thích ứng trong các hệ thống quản lý học tập để mang lại nhiều cơ hội học tập và rèn luyện cho đối tượng người học tại cơ sở giáo dục. Việc xây dựng, quản lý, bảo trì và nâng cấp hệ thống trắc nghiệm thích ứng có thể đặt ra nhiều thách thức, nhưng những ưu điểm và ứng dụng của trắc nghiệm thích ứng hứa hẹn sẽ mang lại những lợi ích không hề nhỏ cho người học, từ đó nâng cao chất lượng đào tạo nói chung của cơ sở giáo dục.

3.3. Khuyến nghị với nhóm chuyên gia phát triển hệ thống

Việc phát triển hệ thống cần thực hiện liên tục, đòi hỏi công sức và nỗ lực không ngừng từ nhóm chuyên gia phát triển hệ thống. Dựa vào những kết quả khả quan đã ghi nhận được với các nội dung kiểm tra đánh giá hiện tại, nhóm nghiên cứu cần phát huy ưu điểm của trắc nghiệm thích ứng, đồng thời phát triển các nội dung kiểm tra đánh giá khác, phù hợp với các mục đích và đối tượng khác nhau, mang trắc nghiệm thích ứng tiếp cận với đông đảo người học ở Việt Nam. Việc áp dụng các mô hình IRT hai hay ba tham số cũng như

việc báo cáo kết quả phù hợp với mục đích đánh giá của người học và chương trình học cũng cần được xem xét một cách nghiêm túc, phối hợp các bên liên quan để mang lại hiệu quả kiểm tra đánh giá hữu ích nhất của hệ thống trong tương lai gần.

Cho dù các phản hồi nhận được đến thời điểm hiện tại là rất tích cực, nhóm chuyên gia phát triển hệ thống vẫn cần đối mặt với những thách thức về việc quản lý và duy trì hệ thống, việc chỉnh sửa các thuật toán thích ứng, nâng cấp các tính năng của hệ thống để mang lại những trải nghiệm trọn vẹn hơn nữa, hỗ trợ quá trình sử dụng của người dùng.

3.4. Khuyến nghị với các nhà nghiên cứu

Việc áp dụng trắc nghiệm thích ứng dù được đón nhận tích cực nhưng vẫn còn khá mới mẻ trong bối cảnh giáo dục ở Việt Nam. Việc xây dựng ngân hàng câu hỏi chuẩn hóa, áp dụng trắc nghiệm thích ứng trong các mục đích khác nhau, từ thực tiễn đo lường, đánh giá trong giáo dục đến thực tiễn dạy và học, đang để lại những khoảng trống nghiên cứu, kêu gọi những nỗ lực nghiên cứu của các học giả. Những nghiên cứu về thuật toán, về ngân hàng câu hỏi, những nghiên cứu về việc phát triển nội dung, tính năng hay xác trị hứa hẹn sẽ mang lại những minh chứng thuyết phục cho việc áp dụng rộng rãi hơn của trắc nghiệm thích ứng cũng như áp dụng rộng rãi hơn các công nghệ tiên tiến trong thực tiễn giáo dục, từ đó đóng góp vào việc nâng cao chất lượng giáo dục và đào tạo của đất nước.

3.5. Khuyến nghị với các cơ quan quản lý giáo dục

Trắc nghiệm thích ứng là một xu hướng phát triển hiện đại trong bối cảnh giáo dục ở Việt Nam, tuy nhiên vẫn còn gặp nhiều thách thức để có thể áp dụng rộng rãi trong giáo dục và đào tạo, cụ thể trong lĩnh vực kiểm tra đánh giá. Nhìn nhận được những cơ hội và thách thức của việc áp dụng trắc nghiệm thích ứng

trong đào tạo và khảo thí, các cơ quan quản lý giáo dục cần có những bước đi nhạy bén và phù hợp để hỗ trợ việc áp dụng trắc nghiệm thích ứng, từ đó giúp nâng cao chất lượng giáo dục và đào tạo, cũng mang lại lợi ích cho các bên liên quan trong lĩnh vực giáo dục. Một số khuyến nghị bao gồm xây dựng, điều chỉnh các chính sách và quy định tạo điều kiện cho việc áp dụng trắc nghiệm thích ứng; tài trợ các nghiên cứu về việc phát triển hệ thống chất lượng bền vững cũng như việc đánh giá hiệu quả của trắc nghiệm thích ứng trong dạy và học cũng như trong kiểm tra đánh giá; đào tạo nhân lực để phát triển, vận hành và bảo trì các hệ thống trắc nghiệm thích ứng; cung cấp hỗ trợ kỹ thuật và tư vấn cho các cơ sở giáo dục và giáo viên về cách triển khai và sử dụng trắc nghiệm thích ứng trong dạy và học cũng như trong kiểm tra đánh giá; và tổ chức các sự kiện, hội thảo, hay các diễn đàn để các cơ sở giáo dục có thể chia sẻ kinh nghiệm và thành công trong việc sử dụng trắc nghiệm thích ứng.

4. Hạn chế của luận án và đề xuất nghiên cứu tiếp theo

Luận án vẫn có những hạn chế nhất định trong quá trình nghiên cứu. Thứ nhất, luận án chỉ dừng lại ở mở rộng nội dung đánh giá của hệ thống trắc nghiệm thích ứng, chưa hướng tới các thuật toán của hệ thống trắc nghiệm thích ứng. Để tạo ra một hệ thống trắc nghiệm thích ứng hiệu quả, các thuật toán đóng vai trò không hề nhỏ, việc lựa chọn thuật toán sử dụng trong hệ thống có thể ảnh hưởng đến hiệu quả đo lường đánh giá của một nội dung cụ thể. Vì vậy, các nghiên cứu trong tương lai có thể xem xét đến việc thử nghiệm các thay đổi về thuật toán, như thuật toán lựa chọn câu hỏi đầu tiên hay thuật toán dừng bài kiểm tra để có thể tìm ra những thuật toán hiệu quả nhất phục vụ cho mục đích đó lường và đánh giá cụ thể.

Thứ hai, số lượng thí sinh tham gia nghiên cứu còn hạn chế. Các nghiên cứu trong tương lai có thể xem xét mở rộng quy mô mẫu và điều tra với đối tượng người học ngoại ngữ tiếng Anh ở Việt Nam với nhiều trình độ ngôn ngữ

và bối cảnh học tập khác nhau. Việc mở rộng cỡ mẫu trong thử nghiệm cũng sẽ dẫn đến việc tăng cỡ mẫu trong khảo sát và phỏng vấn để có những góc nhìn đa dạng hơn từ nhiều đối tượng hơn.

Ngoài ra, nghiên cứu này để lại khoảng trống cho các nghiên cứu xác trị trong tương lai bằng cách sử dụng các mô hình phân tích hai hoặc ba tham số và các gói phân tích khác nhau trong R để mang lại kết quả đa chiều và hữu ích hơn.

Cuối cùng, các nghiên cứu trong tương lai có thể hướng tới xây dựng ngân hàng câu hỏi với các nội dung khác hoặc áp dụng trắc nghiệm thích ứng để phục vụ kiểm tra và đánh giá thường xuyên, hỗ trợ thực hành dạy, học và nghiên cứu giáo dục, cũng như áp dụng cho sơ khảo trình độ ngoại ngữ trong các kỳ thi lấy chứng chỉ ngoại ngữ ở Việt Nam.

DANH MỤC CÔNG TRÌNH NGHIÊN CỨU

LIÊN QUAN ĐẾN LUẬN ÁN

1. **Bùi Thị Kim Phượng**, Nguyễn Quý Thanh và Lê Thái Hưng, “*Computerized Adaptive Testing in Language Education: Opportunities and Challenges in Era 4.0*”, Kỷ yếu Hội thảo Hafpes, tr. 264-274, 2021.
2. **Bùi Thị Kim Phượng**, “*A Review on Vocabulary Tests of High Frequency English Words*”, VNU Journal of Science: Education Research, Vol. 38 (3), tr. 51-60, 2022.
3. **Bùi Thị Kim Phượng**, Nguyễn Quý Thanh và Lê Thái Hưng, “*The Development of a Vietnamese-English Bilingual Version of the New General Service List Test*”, Kỷ yếu Hội thảo Hafpes 2, tr. 502-509, 2022.
4. **Bùi Thị Kim Phượng**, Nguyễn Quý Thanh và Lê Thái Hưng, “*Assessing the Quality of a Newly Designed Vocabulary Test for Vietnamese EFL Learners: A Rasch-based Analysis*”, Vietnam Journal of Education, Vol. 7, tr. 63-73, 2023.
5. Nguyễn Thái Hà, **Bùi Thị Kim Phượng** và Lê Thái Hưng, “*Phát triển ngân hàng câu hỏi trắc nghiệm thích ứng đánh giá từ vựng tiếng Anh thông dụng: áp dụng IRT và phương pháp cân bằng đề*”, Tạp chí Giáo dục, Vol. 23 (19), tr. 8-14, 2023.
6. **Bùi Thị Kim Phượng**, Nguyễn Quý Thanh và Lê Thái Hưng, “*Perceptions of a Computerized Adaptive Test of English Vocabulary: An Online Survey with Vietnamese Test-takers*”, Kỷ yếu Hội thảo Hafpes 3, tr. 481-491, 2023.
7. **Bùi Thị Kim Phượng**, “*Phỏng vấn thí sinh về bài trắc nghiệm thích ứng từ vựng tiếp nhận tiếng Anh*”, Tạp chí Giáo dục và Xã hội, Số đặc biệt Tháng 10, tr.104-108, 2023.
8. **Bùi Thị Kim Phượng**, Lê Thái Hưng, Trần Thị Thu Anh và Trần Thị Tú Anh, “*Kiểm tra, đánh giá kiến thức từ vựng của sinh viên đại học với danh sách từ vựng tiếng Anh thông dụng mới*”, Tạp Chí Giáo dục, Vol.24 (12), tr. 24–28, 2024.

TÀI LIỆU THAM KHẢO

Tài liệu tiếng Việt:

- Bộ Giáo dục và Đào tạo (2014). *Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam* được ban hành kèm theo Thông tư số 01/2014/TT-BGDĐT ngày 24 tháng 01 năm 2014 của Bộ trưởng Bộ Giáo dục và Đào tạo.
- Bộ Giáo dục và Đào tạo (2018). *Chương trình Giáo dục phổ thông môn Tiếng Anh* được ban hành kèm theo Thông tư số 32/2018/TT-BGDĐT ngày 26 tháng 12 năm 2018 của Bộ trưởng Bộ Giáo dục và Đào tạo.
- Lâm Quang Thiệp (2010). *Đo lường trong giáo dục, lý thuyết và ứng dụng*. Nhà xuất bản Đại học Quốc gia Hà Nội.
- Lê Thái Hưng & Nguyễn Thái Hà (2020). *Hệ thống Trắc nghiệm thích ứng phiên bản UEd-CAT 1.0: Đánh giá năng lực thông minh trong kĩ nguyên số*. <http://qm.education.vnu.edu.vn/he-thong-trac-nghiem-thich-ung-uedcat10>
- Lê Thái Hưng & Nguyễn Thái Hà (2021). Xu thế kiểm tra, đánh giá năng lực người học trên nền tảng công nghệ. *Tạp chí Khoa học Giáo dục Việt Nam*, 42, 1-6.
- Lê Thái Hưng, Trần Thị Hoa, Đặng Thị Mây & Hoàng Lan Hương (2019). Phát triển ngân hàng trắc nghiệm thích ứng để đánh giá năng lực đọc hiểu môn Ngữ văn của học sinh lớp 10 trung học phổ thông. *Tạp chí Khoa học Giáo dục Việt Nam*, 24, 54-59.
- Nguyễn Quỳnh Giang & Lê Thái Hưng (2018). Mô phỏng một bài kiểm tra trắc nghiệm thích ứng thông qua phần mềm R. *Tạp chí Khoa học Giáo dục Việt Nam*, 11, 6-11.
- Nguyễn Thái Hà, Vũ Trọng Lương, Lê Thái Hưng & Phạm Văn Hoàng (2021). Thiết kế câu hỏi trắc nghiệm thích ứng nhằm đánh giá năng lực toán học của học sinh lớp 12. *Tạp chí Giáo dục*, 508(2), 33–40.

Tài liệu tiếng Anh:

- Alderson, C. (2007). Computer-adaptive language testing. In S. Granger (Ed.), *Optimizing the role of language in Technology-Enhanced Learning* (pp. 1–4), NOE-Kaleidoscope.
- Apostolou, B., Blue, M. A., & Daigle, R. J. (2009). Student perceptions about computerized testing in introductory managerial accounting. *Journal of Accounting Education*, 27(2), 59-70.
- Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The new computer adaptive test of size and strength (CATS): Development and validation. *Language Assessment Quarterly*, 16(3), 345-368.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.
- Bayona-Ore, L., Zavala, R. F., & Cruz, M. L. (2018, October). Expert opinion process: applications in education. In *Proceedings of the 10th International Conference on Education Technology and Computers* (pp. 172-176). Association for Computing Machinery.
- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language testing*, 27(1), 101-118.
- Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language testing*, 16(2), 131-162.
- Brenzel, J., & Settles, B. (2017). The Duolingo English test—Design, validity, and value. *DET Whitepaper (Short)*. <https://doi.org/10.46999/lyqs3238>

- Brown, A., & Iwashita, N. (1996). Language background and item difficulty: The development of a computer-adaptive test of Japanese. *System*, 24, 199–206.
- Brown, J. D. (1997). Computers in language testing: Present research and some future directions. *Language Learning & Technology*, 1(1), 44-59.
- Browne, C. (2013). The new general service list: Celebrating 60 years of vocabulary learning. *The Language Teacher*, 37(4), 13-16.
- Burston, J. (1995). Practical design and implementation considerations of a computer adaptive foreign language test: The Monash/Melbourne French CAT. *Calico Journal*, 26-46.
- Burston, J., & Neophytou, M. (2014). Lessons Learned in Designing and Implementing a Computer-Adaptive Test for English. *The EuroCALL Review*, 22(2), 19–25. <https://doi.org/10.4995/eurocall.2014.3632>
- Canale, M. (1986). The promise and threat of computerized adaptive assessment of reading comprehension. In C. Stansfield (ed.), *Technology and language testing* (pp. 30-45). TESOL Publications.
- Carlson, J. (2020). *Introduction to item response theory models and applications*. Routledge.
- Chalhoub-Deville, M. (1999). *Issues in computer-adaptive testing of reading proficiency*. University of Cambridge Local Examinations Syndicate.
- Chalhoub-Deville, M. (2010). Technology in standardized language assessments. In R. Kaplan (Ed.), *Handbook of applied linguistics* (2nd ed., pp. 511–26). Oxford University Press.
- Chalhoub–Deville, M., & Deville, C. (1999). Computer adaptive testing in second language contexts. *Annual Review of Applied Linguistics*, 19, 273–299. <https://doi.org/10.1017/s0267190599190147>

- Chapelle, C. A. (1999). Validity in language assessment. *Annual review of applied linguistics*, 19, 254-272.
- Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge University Press.
<https://doi.org/10.1017/cbo9780511733116.002>
- Chapelle, C. A., & Voss, E. (2008). Utilizing technology in language assessment. *Encyclopedia of language and education*, 7, 123-134.
- Chen, J., & Wang, L. (2010, October). Computerized adaptive testing: A new trend in language testing. In *2010 International Conference on Artificial Intelligence and Education (ICAIE)* (pp. 725-728). IEEE.
- Choi, I.-C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, 20, 295–320. <https://doi.org/10.1191/0265532203lt258oa>
- Choi, Y., & McClenen, C. (2020). Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Applied Sciences*, 10(22), 81-96.
<https://doi.org/10.3390/app10228196>
- Coxhead, A., Nation, P., & Sim, D. (2015). Measuring the vocabulary size of native speakers of English in New Zealand secondary schools. *New Zealand Journal of Educational Studies*, 50(1), 121-135.
- Creswell, J & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. SAGE.
- Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches*. Sage Publications.
- Culligan, B. (2015). A comparison of three test formats to assess word difficulty. *Language Testing*, 32(4), 503-520.

- Dandonoli, P. (1989). The ACTFL Computerized Adaptive Test of Foreign Language Reading Proficiency. In F. Smith (Ed.), *Modern Technology in Foreign Language Education: Application and Projects* (pp. 291-300). National Textbook Company.
- Dang, T. N. Y. (2020). Vietnamese non-English major EFL university students' receptive knowledge of the most frequent English words. *VNU Journal of Foreign Studies*, 36(3), 1-11. <https://doi.org/10.25073/2525-2445/vnufs.4553>
- David, A., & Ida, M. (2019). *A Course in Rasch Measurement Theory: Measuring in the Educational, Social and Health Sciences*. Springer.
- Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teachers guide to writing and using language test specifications*. Yale University Press.
- Dunkel, P. (1999). Research and development of a computer-adaptive test of listening comprehension in the less-commonly taught language Hausa. In M. Chalhoub-Deville (Ed.), *Issues in computer-adaptive testing of reading proficiency* (pp. 91-121). University of Cambridge Local Examinations Syndicate.
- Elgort, I. (2013). Effects of L1 definitions and cognate status of test items on the Vocabulary Size Test. *Language Testing*, 30(2), 253-272.
- Fowler Jr, F. J. (2014). *Survey research methods (5th ed.)*. Sage publications.
- Gawliczek, P., Krykun, V., Tarasenko, N., Tyshchenko, M., & Shapran, O. (2021). Computer adaptive language testing according to NATO STANAG 6001 Requirements. *Advanced Education*, 8(17), 19–26. <https://doi.org/10.20535/2410-8286.225018>

- Giouroglou, H., & Economides, A. (2003). *Cognitive CAT in foreign language assessment*. In *Proceedings of Eleventh International PEG Conference, Powerful ICT Tools for Learning and Teaching*. PEG.
- Giouroglou, H., & Economides, A. (2005). *An implemented theoretical framework for a common European foreign language adaptive assessment*. In *Proceedings of 3rd International Conference on Open and Distance Learning*. Greek Open University.
- Godfroid, A. (2019). Sensitive Measures of Vocabulary Knowledge and Processing: Expanding Nation's Framework 1. In *The Routledge handbook of vocabulary studies* (pp. 433-453). Routledge.
- Gyllstad, H. (2019). Measuring knowledge of multiword items. In *The Routledge handbook of vocabulary studies* (pp. 387-405). Routledge.
- He, L., & Min, S. (2017). Development and validation of a computer adaptive EFL test. *Language Assessment Quarterly*, 14(2), 160–176. <https://doi.org/10.1080/15434303.2016.1162793>
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Newbury House.
- Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *Journal of Chiropractic Education*, 33(2), 151-163.
- Hussain, R., Zahid, M. S., & Alam, Y. (2021). Ensuring Viable Solutions in Resource-Constrained Language Teaching Settings: ESL Learners Perceptions about Computer-Assisted Language Testing (CALT). *Sustainable Business and Society in Emerging Economies*, 3(4), 509-519.
- Iriste, S., & Katane, I. (2018). Expertise as a research method in education. *Rural Environment. Education. Personality*, 11(2), 74-80.

- Jamieson, J. (2005). Trends in computer-based second language assessment. *Annual Review of Applied Linguistics*, 25, 228-242.
- Kamimoto, T. (2014). Local item dependence on the vocabulary levels test revisited. *Vocabulary Learning and Instruction*, 3(2), 56-68.
- Karami, H. (2012). The development and validation of a bilingual version of the Vocabulary Size Test. *RELC journal*, 43(1), 53-67.
- Kaya-Carton, E., Carton, A. S., & Dandonoli, P. (1991). Developing a computer-adaptive test of French reading proficiency. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues and practice* (pp. 259-84). New York: Newbury House.
- Keng, L. (2008). *A comparison of the performance of testlet-based computer adaptive tests and multistage tests*. The University of Texas at Austin.
- Khoshsima, H., & Toroujeni, S. M. H. (2017). Computer adaptive testing (Cat) design; testing algorithm and administration mode investigation. *European Journal of Education Studies*.
<http://dx.doi.org/10.46827/ejes.v0i0.725>
- Kimura, T. (2017). The impacts of computer adaptive testing from a variety of perspectives. *Journal of educational evaluation for health professions*, 14.
- Koizumi, R. (2015). Second language vocabulary assessment studies: Validity evidence and future directions. *Vocabulary Learning and Instruction*, 4(1), 36-46.
- Kremmel, B. (2019). Measuring Vocabulary Learning Progress. In *The Routledge handbook of vocabulary studies* (pp. 406-418). Routledge.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

- Kustiyahningsih, Y., & Cahyani, A. D. (2013). Computerized adaptive test based on item response theory in e-learning system. *International Journal of Computer Applications*, 81(6). <https://doi.org/10.5120/14014-2022>
- Lange, D. L. (1990). 'Priority Issues in the Assessment of Communicative Language Abilities.' *Foreign Language Annals*, 23, 403-407.
- Larson, J. W. & Madsen, H. S. (1985). Computer-adaptive language testing: Moving beyond computer-assisted testing. *CALICO Journal*, 2(3), 32–36. <https://doi.org/10.1558/cj.v2i3.32-37>
- Larson, J. W. (1987) Computerized adaptive language testing: A Spanish placement exam. In K. M. Bailey, T. L. Dale & R. T. Clifford (Eds.), *Language Testing Research*, pp. 1-10. Defense Language Institute.
- Larson, J. W. (1989) S-CAPE: A Spanish computerized adaptive placement exam. In Smith, W. F. (Ed.), *Modern Technology in Foreign Language Education: Applications and Projects*, pp. 277-89. National Textbook Company.
- Larson, J. W. (1998). An argument for computer adaptive language testing. *Multimedia-Assisted Language Learning*, 1(1), 9-24.
- Laufer, B. (2013). Lexical thresholds for reading comprehension: What they are and how they can be used for teaching purposes. *Tesol Quarterly*, 47(4), 867–872. <https://doi.org/10.1002/tesq.140>
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language learning*, 54(3), 399-436.
- Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language testing*, 16(1), 33-51.

- Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge?. *Language testing*, 21(2), 202-226.
- Laurier, M. (1999). The development of an adaptive test for placement in French. In M. Chalhoub-Deville (Ed.), *Issues in computer-adaptive testing of reading comprehension* (pp. 122–135). Cambridge University Press.
- Laurier, M. (2000). Can computerized testing be authentic? *ReCALL*, 12(1), 93–104. <https://doi.org/10.1017/s0958344000001014>
- Le, T. C. N, & Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC journal*, 42(1), 86–99. <https://doi.org/10.26686/wgtn.12552149>
- Le, T. H. & Nguyen T. H. (2021). *Experimental Research and Application of Computerized Adaptive Tests to Assess Learners' Competencies*. 2021 3rd International Conference on Computer Science and Technologies in Education. CSTE. <https://doi.org/10.1109/cste53634.2021.00021>
- Le, T. H., Tang, T. T, Tran L. A., Nguyen T. D., Nguyen P. A & Nguyen T. Q. G. (2019). Developing Computerized Adaptive Testing: An Experimental Research on Assessing the Mathematical Ability of 10th Graders. *VNU Journal Of Science: Education Research*, 35(4), 49–63. <https://doi.org/10.25073/2588-1159/vnuer.4301>
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior research methods*, 44(2), 325-343.
- Lin, L. H., & Morrison, B. (2010). The impact of the medium of instruction in Hong Kong secondary schools on tertiary students' vocabulary. *Journal*

- of *English for Academic Purposes*, 9(4), 255–266.
<https://doi.org/10.1016/j.jeap.2010.09.002>
- Liu, X. (2019). *Optimizing design of incorporating off-grade items for constrained computerized adaptive testing in K-12 assessment* (Doctoral dissertation, The University of Iowa).
- Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 179–193.
<https://doi.org/10.1111/j.1745-3984.1980.tb00825.x>
- Madsen, H. (1991). Computer-adaptive testing of listening and reading comprehension. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues and practice* (pp. 237-257). Newbury House.
- Malabonga, V. (2000). *Trends in foreign language assessment: The computerized oral proficiency instrument*. NCLRC Newsletter.
- Malabonga, V., & Kenyon, D. (1999). Multimedia computer technology and performance-based language testing: a demonstration of the computerized oral proficiency instrument. In M. B. Olsen (Ed.), *Computer mediated language assessment and evaluation in natural language processing* (pp. 16-23). Association for Computational Linguistics.
- McLean, S., & Kramer, B. (2015). The creation of a new vocabulary levels test. *Shiken*, 19(2), 1-11.
- McLean, S., & Kramer, B. (2016). The development of a Japanese bilingual version of the New Vocabulary Levels Test. *Vocabulary Education and Research Bulletin*, 5(1), 2-5.
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741-760.
- Meara, P. (1992). *EFL vocabulary tests*. Centre for Applied Language Studies.

- Meara, P. M. (1987). *Vocabulary in a second language (Vol. 2. Specialized Bibliography 4)*. Centre for Information on Language Teaching and Research (CILT).
- Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, 28(1), 19-30.
- Meunier, L. E. (1994). Computer adaptive language tests (CALT) offer a great potential for functional testing. Yet, Why Don't They?. *CALICO journal*, 11(4), 23–39. <https://doi.org/10.1558/cj.v11i4.23-39>
- Milton, J. & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: do vocabulary tests underestimate the knowledge of some learners. *The Canadian Modern Language Review*, 63(1),127-147.
- Milton, J. (2009). *Measuring second language vocabulary acquisition (Vol. 45)*. Multilingual Matters.
- Mizumoto, A., Sasao, Y., & Webb, S. A. (2019). Developing and evaluating a computerized adaptive testing version of the Word Part Levels Test. *Language Testing*, 36(1), 101–123. <https://doi.org/10.1177/0265532217725776>
- Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle and Heinle.
- Nation, I. S. P. (2001). How many high frequency words are there in English?. *Language, Learning and Literature: Studies Presented to Hakan Ringbom*, 4, 167–181.
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9–13.

- Nation, I. S., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.
- Nation, P. (1983). Testing and teaching vocabulary. *Guidelines*, 5, 12–25.
- Nguyen, C. D. (2021). Lexical Features of Reading Passages in English-language Textbooks for Vietnamese High-school Students: Do they Foster both Content and Vocabulary Gain?. *RELC Journal*, 52(3), 509–522. <https://doi.org/10.1177/0033688219895045>
- Nguyen, L. T. C., & Nation, P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal*, 42(1), 86–99.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298–320. <https://doi.org/10.1177/1362168816639619>
- Nguyen, V. C., & Nguyen P. H. (2020). Analyzing and selecting multiple choice test items based on classical test theory and item response theory. *Ho Chi Minh City University of Education Journal of Science*, 17(10), 1804-1818.
- Nogami, Y., & Hayashi, N. (2010). A Japanese adaptive test of English as a foreign language: Development and operational aspects. In W. J. van der Linden, & C. A. W. Glas (Eds.), *Elements of adaptive testing* (pp. 191–211). New York, NY: Springer.
- Norris, J. (2001). Concerns with computerized adaptive oral proficiency assessment. *Language Learning & Technology*, 5(2), 99-105.
- Nydick, S. W., & Weiss, D. J. (2009). A hybrid simulation procedure for the development of CATs. In *Proceedings of the 2009 GMAC Conference on*

Computerized Adaptive Testing. Retrieved from www. psych. umn. edu/psylabs/CATCentral.

Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *The Modern Language Journal*, 93, 836–847.

Okhotnikova, A., Daminova, J., Muzafarova, A., & Rasskazova, T. (2019). Challenges of designing and administering computer-adaptive tests. In *Proceedings of 13th International Technology, Education and Development Conference. INTED.*
<https://doi.org/10.21125/inted.2019.1383>

Olsen, J. B., Maynes, D. D., Slawson, D., & Ho, K. (1989). Comparisons of paper-administered, computer-administered and computerized adaptive achievement tests. *Journal of Educational Computing Research*, 5(3), 311-326.

Oppl, S., Reisinger, F., Eckmaier, A., & Helm, C. (2017). A flexible online platform for computerized adaptive testing. *International Journal of Educational Technology in Higher Education*, 14(1), 1–21.
<https://doi.org/10.1186/s41239-017-0039-0>

Paribakht, T., & Wesche, M. (1996). Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition: A Rationale for Pedagogy*, pp. 174-200. Cambridge University Press.
<https://doi.org/10.1017/cbo9781139524643.013>

Pathan, M. M. (2012). Computer Assisted Language Testing [CALT]: advantages, implications and limitations. *Research Vistas*, 1(4), 30-45.

- Qian, D. D., & Lin, L. H. (2019). The relationship between vocabulary knowledge and language proficiency. In *The Routledge handbook of vocabulary studies*, pp. 66-80. Routledge.
- Rasskazova, T., Muzafarova, A., Daminova, J., & Okhotnikova, A. (2017). Computerised language assessment: Limitations and Opportunities. *eLearning & Software for Education*, 2. <https://doi.org/10.12753/2066-026x-17-110>
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC journal*, 19(2), 12-25.
- Read, J. (2000). *Assessing vocabulary*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511732942>
- Read, J. (2019). Key issues in measuring vocabulary knowledge. In *The Routledge handbook of vocabulary studies* (pp. 545-560). Routledge.
- Read, J., & Chapelle, C. A. (2001). A framework for second language vocabulary assessment. *Language testing*, 18(1), 1-32.
- Rudner, L. M. (2009). Implementing the graduate management admission test computerized adaptive test. In *Elements of adaptive testing* (pp. 151–165). Springer New York. https://doi.org/10.1007/978-0-387-85461-8_8
- Şahin, A., & Weiss, D. J. (2015). Effects of calibration sample size and item bank size on ability estimation in computerized adaptive testing. *Educational Sciences: Theory & Practice*, 15(6). <https://doi.org/10.12738/estp.2015.6.0102>
- Sasao, Y., & Webb, S. (2017). The Word Part Levels Test. *Language Teaching Research*, 21, 12–30.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. <https://doi.org/10.1057/9780230293977>

- Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951.
<https://doi.org/10.1111/lang.12077>
- Schmitt, N., & Zimmerman, C. B. (2002). Derivative word forms: What do learners know?. *TESOL quarterly*, 36(2), 145–171.
<https://doi.org/10.2307/3588328>
- Schmitt, N., Cobb, T., Horst, M., & Schmitt, D. (2017). How much vocabulary is needed to use English? Replication of van Zeeland & Schmitt (2012), Nation (2006) and Cobb (2007). *Language Teaching*, 50(2), 212–226.
<https://doi.org/10.1017/s0261444815000075>
- Schmitt, N., Nation, P., & Kremmel, B. (2020). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 109-120.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language testing*, 18(1), 55-88.
- Seidman, I. (2019). *Interviewing as qualitative research: A guide for researchers in education and the social sciences (5th ed.)*. Teachers College Press.
- Spoden, C., Frey, A., & Bernhardt, R. (2018). Implementing three CATs within eighteen months. *Journal of Computerized Adaptive Testing*, 6(3).
- Starr-Egger, F. (2001). German computer adaptive language tests—better late than never. *German as a Foreign Language*, 3, 12-30.
- Stevenson, J., & Gross, S. (1991). Use of a computerized adaptive testing model for ESOL/bilingual entry/exit decision making. In P. Dunkel (Ed.),

- Computer-assisted language learning and testing: Research issues and practice*, pp. 237-57. Newbury House.
- Stewart, J. (2009). A comparison of test scores between monolingual and bilingual versions of the Vocabulary Size Test: A pilot study. In *JALT 2008 Conference Proceedings* (pp. 1262-1269). JALT.
- Stoeckel, T., & Bennett, P. (2015). A test of the new General Service List. *Vocabulary Learning and Instruction*, 4(1), 1–8. <https://doi.org/10.7820/vli.v04.1.stoeckel.bennett>
- Stoeckel, T., Ishii, T., & Bennett, P. (2018). A Japanese-English bilingual version of the new general service list test. *JALT Journal*, 40(1), 5–21. <https://doi.org/10.37546/jaltjj40.1-1>
- Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, 43(1), 181-203.
- Stone, E., & Davey, T. (2011). Computer Adaptive Testing for Students with Disabilities: A Review of the Literature. *ETS Research Report Series*, 2011(2).
- Suvorov, R., & Hegelheimer, V. (2013). Computer-assisted language testing. *The companion to language assessment*, 2, 594-613.
- Szabo, G. (2008). *Applying Item Response Theory in Language Test Item Bank Building*. Peter Lang.
- Thompson, N. A., & Weiss, D. A. (2011). A framework for the development of computerized adaptive tests. *Practical Assessment, Research, and Evaluation*, 16(1), 1. <https://doi.org/10.7275/wqzt-9427>
- Thorndike, E. & Lorge, I. (1944). *The teacher's word book of 30,000 words*. Teachers College Press.

- Triantafillou, E., Georgiadou, E., & Economides, A. A. (2008). The design and evaluation of a computerized adaptive test on mobile devices. *Computers & Education, 50*(4), 1319-1330.
- Tseng, W. T. (2016). Measuring English vocabulary size via computerized adaptive testing. *Computers & Education, 97*, 69–85. <https://doi.org/10.1016/j.compedu.2016.02.018>
- Tung, P. (1986). Computerized adaptive testing: Implications for language test developers. In C. W. Stansfield (Ed.). *Technology and language testing* (pp. 9-11). TESOL.
- Vispoel, W. P. (1993). Computerized adaptive and fixed-item versions of the ITED vocabulary subtest. *Education and Psychological Measurement, 53*, 779-788.
- Vispoel, W. P. (1998). Psychometric characteristics of computer-adaptive and self-adaptive vocabulary tests: the role of answer feedback and test anxiety. *Journal of Educational Measurement, 35*, 155-167.
- Vispoel, W. P., Rocklin, T. R., & Wang, T. (1994). Individual differences and test administration procedures: a comparison of fixed-item, computerized adaptive, and self-adapted testing. *Applied Measurement in Education, 7*, 53-79.
- Vu, D. V., & Nguyen, N. C. (2019). *An assessment of vocabulary knowledge of Vietnamese EFL learners*. In *Proceedings of The 20th English in Southeast Asia Conference*. Nanyang Technological University.
- Vu, D. V., & Peters, E. (2021). Vocabulary in English language learning, teaching, and testing in Vietnam: A review. *Education Sciences, 11*(9), 563.

- Wainer, H., & Eignor, D. (2000). Caveats, pitfalls, and unexpected consequences of implementing large-scale computerized testing. In H. Wainer et al. (Eds.), *Computer adaptive testing: A primer (2nd ed.)* (pp. 271–299). Laurence Erlbaum Associates.
- Wang, H. P., Kuo, B. C., Chao, R. C., & Tsai, Y. H. (2012). The development and evaluation of a computerized adaptive testing system for Chinese proficiency-base on CEFR. *Procedia-Social and Behavioral Sciences*, *64*, 34-42.
- Webb, S. A., & Chang, A. C. S. (2012). Second language vocabulary growth. *RELC journal*, *43*(1), 113–126.
<https://doi.org/10.1177/0033688212439367>
- Webb, S., & Piasecki, A. (2018). Re-examining the effects of word writing on vocabulary learning. *ITL-International Journal of Applied Linguistics*, *169*(1), 72-94.
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL-International Journal of Applied Linguistics*, *168*(1), 33-69.
- Wen, K., & Qinghua, L. (2002). The trend of large-scale language testing: CALT [J]. *Foreign Language World*, *2*.
- West, M. (1953). *A general service list of English words*. Longman, Green.
- Wise, S. L. (2014). The utility of adaptive testing in addressing the problem of unmotivated examinees. *Journal of Computerized Adaptive Testing*, *2*(3), 1-17.
- Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Educational Measurement Solutions.

- Wu, M. L., Tam, H. P., & Jen, T. (2016). *Educational Measurement for Applied Researchers Theory into Practice*. Springer.
- Yanagisawa, A., & Webb, S. (2019). Measuring Depth of Vocabulary. *The Routledge handbook of vocabulary studies*, 371.
- Young, R., Shermis, M. D., Brutton, S. R., & Perkins, K. (1996). From conventional to computer-adaptive testing of ESL reading comprehension. *System*, 24(1), 23-40.
- Zareva, A., Schwanenflugel, P., & Nikolova, Y. (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition*, 27(4), 567–595. <https://doi.org/10.1017/s0272263105050254>

Phụ lục 1: Bảng câu hỏi xin ý kiến chuyên gia

Kính chào Thầy/Cô,

Bảng câu hỏi được gửi tới Thầy/Cô để xin ý kiến về đề kiểm tra từ vựng tiếp nhận tiếng Anh được thiết kế phục vụ luận án của NCS Bùi Thị Kim Phượng.

Để giúp nghiên cứu sinh hoàn thiện luận án, kính mong các Thầy/Cô dành ít phút để trả lời các câu hỏi sau đây. Thông tin cá nhân và câu trả lời của Thầy/Cô sẽ được bảo mật và chỉ sử dụng cho mục đích nghiên cứu.

Chân thành cảm ơn Thầy/Cô!

Phần 1: Thông tin chuyên gia

1. Tuổi:
2. Giới tính
 - A. Nam
 - B. Nữ
3. Chứng chỉ tiếng Anh Thầy/Cô sở hữu:
Kết quả cụ thể
4. Bằng cấp cao nhất
 - A. Cử nhân
 - B. Thạc sĩ
 - C. Tiến sĩ
5. Chứng chỉ ra đề/thiết kế bài kiểm tra chuẩn hóa
 - A. Có
 - B. Không
6. Số năm giảng dạy tiếng Anh
7. Mức độ thường xuyên ra đề kiểm tra tiếng Anh (bao gồm cả từ vựng, ngữ pháp, phát âm và các kỹ năng nghe, nói, đọc, viết)
8. Mức độ thường xuyên đánh giá các đề kiểm tra tiếng Anh (bao gồm cả từ vựng, ngữ pháp, phát âm và các kỹ năng nghe, nói, đọc, viết)

Phần 2: Đánh giá đề kiểm tra

Thầy/Cô lựa chọn mức độ đánh giá phù hợp với mỗi tiêu chí trong bảng sau:

Mức độ đánh giá: 5: Rất tốt; 4: Tốt; 3: Khá; 2: Trung bình; 1: Kém; 0: Rất kém

Tiêu chí đánh giá	5: Rất tốt	4: Tốt	3: Khá	2: Trung bình	1: Kém	0: Rất kém
Mục đích kiểm tra rõ ràng						
Nội dung kiểm tra chính xác						
Cấu trúc đề kiểm tra logic						
Đối tượng kiểm tra phù hợp						

Phần 3: Các gợi ý chỉnh sửa về đề kiểm tra và câu hỏi

Trong quá trình đánh giá, Thầy/Cô có phát hiện ra vấn đề gì với câu hỏi hay đề kiểm tra ko? Thầy cô có gợi ý gì để nâng cao chất lượng câu hỏi cũng như đề kiểm tra?

.....

.....

.....

.....

.....

.....

.....

Phụ lục 2: Bảng câu hỏi khảo sát sinh viên

Chào em!

Khảo sát được tiến hành để thu thập ý kiến của sinh viên về bài kiểm tra thích ứng từ vựng tiếng Anh thông dụng. Cô rất mong nhận được ý kiến chia sẻ của em về vấn đề này.

Hãy dành 3-5 phút để trả lời các câu hỏi khảo sát. Thông tin cá nhân và câu trả lời của em sẽ được bảo mật và chỉ sử dụng cho mục đích nghiên cứu.

Chân thành cảm ơn em!

Phần 1: Thông tin cá nhân

Em vui lòng cung cấp một số thông tin bằng cách trả lời các câu hỏi sau. Thông tin của em sẽ được giữ kín và chỉ sử dụng cho mục đích nghiên cứu.

1. Họ tên:
2. MSSV:
3. Giới tính:
4. Em hãy tự đánh giá lượng từ tiếng Anh của em hiện tại:
 - Dưới 1500 từ
 - Từ 1500 đến dưới 2750 từ
 - Từ 2750 đến dưới 3250 từ
 - Từ 3250 đến dưới 3750 từ
 - Từ 3750 đến dưới 4500 từ
 - Từ 4500 trở lên

5. Trước trải nghiệm thích ứng từ vựng tiếng Anh thông dụng, em đã có trải nghiệm với trải nghiệm thích ứng chưa?

- Chưa bao giờ
- 1-3 lần
- Nhiều hơn 3 lần

6. Em sử dụng thiết bị gì để thực hiện trải nghiệm thích ứng từ vựng tiếng Anh thông dụng?

- Máy tính bàn
- Máy tính xách tay
- Máy tính bảng
- Điện thoại thông minh
- Thiết bị khác

Phần 2: Ý kiến của sinh viên về bài kiểm tra thích ứng từ vựng tiếng Anh thông dụng

Em vui lòng lựa chọn mức độ đồng ý của em với các nhận định liên quan đến bài kiểm tra thích ứng từ vựng tiếng Anh thông dụng.

Mức độ thể hiện ý kiến: 4: Hoàn toàn đồng ý; 3: Đồng ý một phần; 2: Phản đối một phần; 1: Hoàn toàn phản đối

	4: Hoàn toàn đồng ý	3: Đồng ý một phần	2: Phản đối một phần	1: Hoàn toàn phản đối
1. Thời gian làm bài kiểm tra ngắn				
2. Bài kiểm tra dễ dàng thực hiện				
3. Bài kiểm tra đánh giá chính xác năng lực thí sinh				
4. Các câu hỏi trắc nghiệm phù hợp với năng lực thí sinh				
5. Thí sinh không thấy căng thẳng khi làm bài				
6. Bài kiểm tra có thể hạn chế gian lận thi cử như thí sinh nhắc bài, hỏi bài, nhớ đề				
7. Thí sinh được thông báo kết quả ngay sau khi làm bài				

8. Hệ thống cung cấp những gợi ý học tập tức thời				
9. Bài kiểm tra là công cụ tự đánh giá hiệu quả				
10. Bài kiểm tra là công cụ hỗ trợ học tập và ôn luyện thường xuyên				

1. Điều gì em thấy HÀI LÒNG NHẤT về trắc nghiệm thích ứng từ vựng tiếng Anh thông dụng?
2. Điều gì làm em thấy CHƯA HÀI LÒNG về trắc nghiệm thích ứng từ vựng tiếng Anh thông dụng?
3. Em có mong muốn trải nghiệm trắc nghiệm thích ứng với các nội dung khác?
 - Rất mong muốn
 - Mong muốn
 - Không mong muốn
 - Rất không mong muốn

Phụ lục 3: Bộ câu hỏi phỏng vấn sinh viên

1. Thông tin cá nhân

- Giới thiệu về bản thân (sv năm nhất, ngành gì, quê quán)
- Việc học tiếng Anh: có thích học không? Học thế nào? E ngại gì? Tự tin gì?
- Đánh giá trình độ tiếng Anh (TOEIC?), lượng từ vựng của em (ước lượng bằng cách nào)

2. Trải nghiệm với CAT

- Em biết gì về trắc nghiệm thích ứng?
- Em có trải nghiệm gì về CAT chưa, kỳ vọng/mong muốn gì?
- Sử dụng thiết bị gì? Có ảnh hưởng đến trải nghiệm CAT của em không?

3. Em có quan điểm thế nào về các tính năng của CAT từ vựng tiếng Anh thông dụng?

1. Thời gian làm bài kiểm tra ngắn
2. Bài kiểm tra dễ dàng thực hiện
3. Bài kiểm tra đánh giá chính xác năng lực thí sinh
4. Các câu hỏi trắc nghiệm phù hợp với năng lực thí sinh
5. Thí sinh không thấy căng thẳng khi làm bài
6. Bài kiểm tra có thể hạn chế gian lận thi cử như thí sinh nhắc bài, hỏi bài, nhớ đề
7. Thí sinh được thông báo kết quả ngay sau khi làm bài
8. Hệ thống cung cấp những gợi ý học tập tức thời
9. Bài kiểm tra là công cụ tự đánh giá hiệu quả
10. Bài kiểm tra là công cụ hỗ trợ học tập và ôn luyện thường xuyên

4. Em có kết luận chung gì về CAT từ vựng tiếng Anh thông dụng?

1. Điều gì em thấy HÀI LÒNG NHẤT về trắc nghiệm thích ứng từ vựng tiếng Anh thông dụng?
2. Điều gì làm em thấy CHƯA HÀI LÒNG về trắc nghiệm thích ứng từ vựng tiếng Anh thông dụng?
3. Em có mong muốn trải nghiệm trắc nghiệm thích ứng với các nội dung khác? Hay có đề xuất gì với các hệ thống trắc nghiệm thích ứng đang và sắp được xây dựng trong tương lai?

Phụ lục 4: Đề kiểm tra song ngữ từ vựng tiếp nhận tiếng Anh

Bạn hãy chọn đáp án có nghĩa của từ in đậm trong câu cho trước.

1. charge: They are the **charges**.
 - a. những điều quan trọng cần suy nghĩ
 - b. các khoản phí cho một dịch vụ
 - c. những điều tốt
 - d. các lý do
2. case: This is a good **case**.
 - a. nơi học tập
 - b. cách thức hoạt động
 - c. trường hợp
 - d. kế hoạch tương lai
3. different: They are **different**.
 - a. dễ thấy
 - b. to lớn
 - c. rất khó
 - d. khác biệt
4. room: Where is the **room**?
 - a. quyển sách
 - b. xe ô tô
 - c. cửa hàng
 - d. phòng
5. lead: I will **lead** you.
 - a. đưa bạn đến một nơi
 - b. gặp bạn
 - c. cho phép bạn
 - d. đưa một cái gì đó
6. policy: That is a good **policy**.
 - a. loại trường học
 - b. câu chuyện
 - c. nơi đến thăm
 - d. chính sách
7. rise: They will **rise** next week.
 - a. tăng lên
 - b. thay đổi
 - c. trở nên tốt hơn
 - d. hoàn thành
8. sure: I am **sure**.
 - a. trẻ
 - b. sớm
 - c. chắc chắn
 - d. mới
9. health: **Health** is important.
 - a. học tập
 - b. sức khỏe
 - c. luyện tập
 - d. nhận sự giúp đỡ
10. expect: I **expected** this.
 - a. dự đoán
 - b. tuyên bố
 - c. đặt đề
 - d. đưa đến nơi
11. include: We are **including** it.
 - a. trả tiền
 - b. thay đổi
 - c. thêm vào
 - d. đọc hiểu
12. building: Where is the **building**?
 - a. đồng nghiệp
 - b. đường phố
 - c. phần giữa
 - d. tòa nhà
13. true: that is **true**.
 - a. chính xác
 - b. khác nhau
 - c. thú vị
 - d. tự nhiên
14. teacher: They are **teachers**.
 - a. phụ huynh
 - b. giáo viên
 - c. lãnh đạo
 - d. thanh niên

15. well: You did that **well**.

- a. nhanh
- b. tốt
- c. một mình
- d. thường xuyên

16. return: Please **return** it.

- a. bàn luận
- b. bán đi
- c. trưng bày
- d. gửi trả

17. result: We had the same **results**.

- a. câu hỏi
- b. suy nghĩ
- c. nguyên tắc
- d. kết quả

18. among: He was **among** them.

- a. sau
- b. trước
- c. ở giữa
- d. không xa

19. consider: She **considered** it.

- a. không tìm thấy
- b. cần
- c. cân nhắc
- d. chọn

20. approach: We like your **approach**.

- a. cách tiếp cận
- b. một phần sách
- c. nhà đất
- d. thông tin

21. observe: We **observed** the people.

- a. đến thăm
- b. nói chuyện với
- c. chỉ dẫn
- d. quan sát

22. variety: I like the **variety** here.

- a. điễm tốt
- b. hàng hóa
- c. sự đa dạng
- d. cách mọi người nghĩ

23. fly: We will **fly** there.

- a. đi nhanh
- b. làm bài kiểm tra
- c. tìm một cái gì đó
- d. đi máy bay

24. significant: This is **significant**.

- a. quan trọng
- b. cơ bản
- c. không phổ biến
- d. tự nhiên

25. track: Where is the **track**?

- a. đường đua
- b. điễm xuất phát
- c. xe đua
- d. trọng tài

26. instance: This, for **instance**, was a good idea.

- a. hợp tác
- b. một chóc lát
- c. sức khỏe
- d. ví dụ

27. solution: This is one **solution**.

- a. giải pháp
- b. ý tưởng để cân nhắc
- c. kế hoạch chi tiêu
- d. thành phần

28. clean: It was **clean**.

- a. nguy hiểm
- b. sạch sẽ
- c. rất tốt
- d. rất nhỏ

29. pressure: Put **pressure** on this.
- nước
 - áp lực
 - hơi lạnh
 - vỏ bọc
30. cross: Please make a **cross**.
- hình giống như một chữ X
 - loại đồ ăn
 - luật lệ
 - nhóm từ
31. benefit: These are the **benefits**.
- chi phí
 - các vấn đề
 - ví dụ
 - lợi ích
32. straight: I went **straight** home.
- vượt ra ngoài
 - thẳng đường
 - xung quanh
 - một cách dễ dàng
33. despite: **Despite** the beginning, the speech was good.
- thêm vào
 - bởi vì
 - dựa theo
 - bất chấp
34. bank: I like this **bank**.
- khoảng thời gian
 - nhóm công nhân
 - ngân hàng
 - thành phần
35. examine: She wants to **examine** this.
- hoàn thành
 - giải thích ý nghĩa
 - kiểm tra
 - đến thăm
36. mix: We **mixed** them.
- trình bày
 - giải quyết
 - trộn lẫn
 - gây rắc rối
37. generation: They are the next **generation**.
- câu hỏi cần trả lời
 - kế hoạch tương lai
 - thế hệ
 - công nhân
38. debate: We had a **debate**.
- bài kiểm tra
 - đề cử lãnh đạo
 - kết quả tệ
 - cuộc tranh luận
39. deep: It is **deep**.
- đơn giản
 - khó mang theo
 - được ưa chuộng
 - sâu
40. financial: The problem was **financial**.
- liên quan tài chính
 - liên quan giáo dục
 - khác biệt
 - đang xảy ra
41. ice: Look at the **ice**.
- ngôi nhà nhỏ
 - đá băng
 - phụ tùng xe hơi
 - cây xanh
42. hall: This is an interesting **hall**.
- hội trường
 - loại động vật
 - âm thanh
 - câu trả lời

43. extra: These are **extra**. a.
 vật phẩm y tế
 b. phần phụ thêm
 c. cần thiết
 d. dễ dàng nhận thấy
44. mechanism: This **mechanism** works well.
 a. nhóm người
 b. bản lề
 c. cơ chế
 d. trợ lý lãnh đạo
45. mail: Do they have any **mail**?
 a. nơi ở
 b. tiền bạc
 c. thư từ
 d. quần áo ấm
46. decrease: Next year, it will **decrease**.
 a. đến với nhau
 b. giảm bớt
 c. di chuyển rất nhanh
 d. di chuyển đến nhiều nơi khác nhau
47. extent: She did not know the **extent** of it.
 a. thời gian
 b. màu sắc
 c. quy mô
 d. nguyên nhân
48. progress: Their **progress** was good.
 a. thái độ
 b. khởi đầu
 c. sự tiến bộ
 d. kết quả
49. guarantee: I **guarantee** this.
 a. bảo đảm
 b. dự đoán
 c. cần
 d. có
50. double: It will become **double**.
 a. nhiều gấp hai lần
 b. rất nhỏ
 c. tệ hơn
 d. thường tình
51. factory: This is a new **factory**.
 a. nhà máy
 b. loại sản phẩm
 c. nhóm đồ vật
 d. luật lệ
52. label: Please look at the **label**.
 a. nhãn/ nhãn hiệu
 b. cuốn sách lớn
 c. thay đổi quy tắc
 d. ảnh chân dung
53. breath: My **breath** was fast.
 a. hơi thở
 b. việc di chuyển
 c. việc định giờ
 d. hiểu biết
55. global: This is a **global** problem.
 a. nhỏ
 b. nghiêm trọng
 c. toàn cầu
 d. thuộc địa phương
56. hide: It was **hidden**.
 a. còn mới
 b. bị giấu đi
 c. bị thay đổi
 d. rất chi tiết
57. comfortable: Are you **comfortable**?
 a. thoải mái
 b. muốn thử
 c. có khả năng
 d. ngạc nhiên

58. quote: Can I **quote** you?.

- a. trích dẫn
- b. phê bình
- c. đi cùng
- d. khuyên nhủ

59. sheet: I need a new **sheet**.

- a. điều cần đọc
- b. nơi ở
- c. tờ giấy
- d. đồ uống

60. confirm: They **confirmed** it.

- a. nghi ngờ
- b. khẳng định
- c. đòi hỏi
- d. thử làm

61. thick: It is very **thick**.

- a. khô
- b. quan trọng
- c. xinh đẹp
- d. to/dày

62. ordinary: He is **ordinary**.

- a. khỏe mạnh
- b. thấp lùn
- c. kiêu ngạo
- d. bình thường

63. steal: It has been **stolen**.

- a. đảo trộn
- b. đánh cắp
- c. theo dõi
- d. biến đổi

64. variable: Those are two **variables**.

- a. biến số
- b. mục đích
- c. vật sở hữu
- d. cách thức làm việc

64. uniform: He has his **uniform**.

- a. đồng phục
- b. đề xuất kinh doanh
- c. trở ngại
- d. lý do

65. bone: It is a **bone**.

- a. âm thanh
- b. máy nhỏ
- c. xương
- d. biệt thự

66. occupy: This place is **occupied**.

- a. để trống
- b. không thích
- c. trong một danh sách
- d. đang được sử dụng

67. impose: They **imposed** their ideas.

- a. giữ vững
- b. viết ra
- c. trình bày
- d. áp đặt

68. disappear: She has **disappeared**.

- a. hoàn thành công việc
- b. thay đổi công việc
- c. thể hiện tốt
- d. biến mất

69. curve: This is a **curve**.

- a. nơi cao hơn
- b. chuồng cho động vật
- c. đường tròn
- d. sông nhỏ

70. owe: I **owe** a lot.

- a. có
- b. tiết kiệm
- c. nợ
- d. làm việc

71. justify: We cannot **justify** this.

- a. đánh mạnh
- b. làm theo
- c. trao đổi
- d. chứng minh

72. sink: We started to **sink**.

- a. chìm
- b. thì thâm
- c. cố gắng đồng ý
- d. có một bữa tiệc

73. freeze: It was **frozen**.

- a. đầy màu sắc
- b. bị đóng băng
- c. bí mật
- d. lạ lùng

74. currency: What is the **currency** there?

- a. thời gian
- b. tình huống
- c. loại tiền tệ
- d. vấn đề

75. sad: That is **sad**.

- a. mạnh
- b. lạnh
- c. yên tĩnh
- d. buồn

76. formula: This is a **formula**.

- a. phòng đợi
- b. nhà máy
- c. thành phố
- d. công thức

77. slide: It started to **slide**.

- a. bị vỡ
- b. trượt xuống
- c. tạo lực
- d. to lớn hơn

78. tension: I don't like the **tension** here.

- a. sự căng thẳng
- b. chất lượng của không khí, đất và nước
- c. thông tin về hàng hóa
- d. giá cả hàng hóa

79. shadow: I can see the **shadow**.

- a. bóng/ hình bóng
- b. sự khác biệt
- c. khoảng trống
- d. nơi bắt đầu

80. cooperation: **Cooperation** is a good thing.

- a. làm việc cùng nhau
- b. giúp đỡ người nghèo
- c. nói chuyện với mọi người
- d. nghĩ về mọi người

81. smooth: It was **smooth**.

- a. xa xôi
- b. rất tệ
- c. mạnh mẽ nhất
- d. trơn nhẵn

82. joy: I had a lot of **joy**.

- a. cảm xúc hạnh phúc
- b. cảm giác không biết điều gì đó
- c. thông tin về một cái gì đó
- d. thay đổi

83. motivation: **Motivation** is important.

- a. động lực
- b. sự quyết đoán
- c. năng lực
- d. sức khỏe

84. journal: It is a **journal**.

- a. loại thuốc
- b. quyết định
- c. tạp chí
- d. phát minh

85. confusion: We had some **confusion**.

- a. nhầm lẫn
- b. hành lý
- c. nhu cầu
- d. hội thoại

86. coin: That is my **coin**.

- a. cái nôi
- b. đồng tiền
- c. con dao
- d. cái ghế

87. gradually: I did it **gradually**.

- a. cuối cùng
- b. không có khó khăn
- c. từ từ/dần dần
- d. nhiều lần

88. castle: It is a **castle**.

- a. khẩu súng
- b. phòng mỹ thuật
- c. lâu đài
- d. nhóm người lớn

89. bias: This is **biased**.

- a. đang thay đổi
- b. không mạnh mẽ
- c. trung thực
- d. không công bằng

90. summary: This is a good **summary**.

- a. khởi đầu
- b. lý do để không làm điều gì đó
- c. khoảng thời gian giữa hai việc
- d. tóm tắt

91. diversity: We like the **diversity** there.

- a. ngày lễ
- b. sự đa dạng
- c. kết nối giữa mọi thứ
- d. cách sống

92. edit: It wasn't **edited** very well.

- a. được chỉnh sửa
- b. được thảo luận
- c. được lên kế hoạch
- d. được hiểu rõ

93. comprehensive: It was very **comprehensive**.

- a. toàn diện
- b. ngạc nhiên

c. được yêu thích

d. ngắn hạn

94. excess: There was **excess** food.

- a. được yêu thích
- b. không bình thường
- c. tốt cho sức khỏe
- d. dư thừa

95. throat: Is your **throat** okay?

- a. ghế
- b. quần áo
- c. cổ họng
- d. thức ăn

96. profession: This is my **profession**.

- a. loại văn bản
- b. trang phục
- c. nghề nghiệp
- d. quyết định

97. moderate: The size is **moderate**.

- a. vừa phải
- b. rất lớn
- c. rất nhỏ
- d. không ổn

98. format: This is the **format**.

- a. đồng nghiệp
- b. điều cần học
- c. vấn đề
- d. định dạng

99. accurate: Is this **accurate**?

- a. chính xác
- b. đắt tiền
- c. khó sử dụng
- d. dễ dàng nhận thấy

100. compound: They are **compounds**.

- a. linh hồn
- b. gợi ý trả lời câu hỏi
- c. tư tưởng
- d. hỗn hợp

Phụ lục 5: Kết quả đánh giá định tính 7 đề kiểm tra từ các chuyên gia
(CG: chuyên gia)

Đề	Mục đích kiểm tra rõ ràng			Nội dung kiểm tra chính xác		
	CG1	CG3	CG 5	CG 1	CG 3	CG 5
1	5	4	5	5	4	5
	CG 2	CG 4	CG 6	CG 2	CG 4	CG 6
2	4	5	4	5	5	5
	CG 1	CG 3	CG 6	CG 1	CG 3	CG 6
3	5	5	5	5	4	4
	CG 2	CG 4	CG 5	CG 2	CG 4	CG 5
4	4	4	5	4	5	4
	CG 1	CG 2	CG 7	CG 1	CG 2	CG 7
5	5	4	5	5	5	4
	CG 3	CG 5	CG 7	CG 3	CG 5	CG 7
6	4	5	5	4	4	4
	CG 4	CG 6	CG 7	CG 4	CG 6	CG 7
7	5	4	5	4	5	5

Đề	Cấu trúc đề kiểm tra logic			Đối tượng kiểm tra phù hợp		
1	CG 1	CG 3	CG 1	CG 3	CG 1	CG 3
	5	5	5	5	5	5
2	CG 2	CG 4	CG 2	CG 4	CG 2	CG 4
	4	4	4	4	4	4
3	CG 1	CG 3	CG 1	CG 3	CG 1	CG 3
	5	5	5	5	5	5
4	CG 2	CG 4	CG 2	CG 4	CG 2	CG 4
	5	4	5	4	5	4
5	CG 1	CG 2	CG 1	CG 2	CG 1	CG 2
	5	5	5	5	5	5
6	CG 3	CG 5	CG 3	CG 5	CG 3	CG 5
	5	5	5	5	5	5
7	CG 4	CG 6	CG 4	CG 6	CG 4	CG 6
	5	5	5	5	5	5

Phụ lục 6: Kết quả phân tích sự phù hợp với mô hình của 7 đề thử nghiệm

Kết quả phân tích sự phù hợp với mô hình của Đề 1

ConQuest: Generalised Item Response Modelling Software Fri Feb 03 20:31 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 1	-4.019	0.513	1.08 (0.78, 1.22)	0.7	1.04 (0.07, 1.93)	0.2		
2 2	-4.321	0.589	0.28 (0.78, 1.22)	-9.1	0.93 (0.00, 2.09)	0.1		
3 3	-4.321	0.589	1.71 (0.78, 1.22)	5.2	1.01 (0.00, 2.09)	0.2		
4 4	-4.740	0.717	0.25 (0.78, 1.22)	-9.7	0.95 (0.00, 2.36)	0.2		
5 5	-1.408	0.203	0.85 (0.78, 1.22)	-1.3	0.97 (0.80, 1.20)	-0.2		
6 6	-2.723	0.300	1.79 (0.78, 1.22)	5.6	1.08 (0.55, 1.45)	0.4		
7 7	0.380	0.178	1.08 (0.78, 1.22)	0.7	1.03 (0.87, 1.13)	0.4		
8 8	-2.334	0.261	0.72 (0.78, 1.22)	-2.7	0.92 (0.64, 1.36)	-0.4		
9 9	-2.916	0.322	0.47 (0.78, 1.22)	-5.8	0.87 (0.50, 1.50)	-0.4		
10 10	-1.449	0.205	0.90 (0.78, 1.22)	-0.9	1.03 (0.79, 1.21)	0.3		
11 11	-4.740	0.717	0.25 (0.78, 1.22)	-9.7	0.95 (0.00, 2.36)	0.2		
12 12	-5.447	1.007	0.32 (0.78, 1.22)	-8.3	1.01 (0.00, 2.95)	0.3		
13 13	0.412	0.179	0.91 (0.78, 1.22)	-0.8	0.94 (0.87, 1.13)	-0.9		
14 14	-3.273	0.372	0.87 (0.78, 1.22)	-1.2	0.96 (0.39, 1.61)	-0.0		
15 15	-1.534	0.208	0.86 (0.78, 1.22)	-1.3	0.97 (0.78, 1.22)	-0.3		
16 16	-2.203	0.250	0.55 (0.78, 1.22)	-4.7	0.85 (0.67, 1.33)	-0.9		
17 17	-1.287	0.198	0.92 (0.78, 1.22)	-0.7	0.93 (0.81, 1.19)	-0.7		
18 18	-0.886	0.185	1.05 (0.78, 1.22)	0.4	1.03 (0.85, 1.15)	0.4		
19 19	-0.784	0.183	0.91 (0.78, 1.22)	-0.8	0.94 (0.86, 1.14)	-0.8		
20 20	-0.365	0.177	0.91 (0.78, 1.22)	-0.8	0.92 (0.88, 1.12)	-1.2		
21 21	-2.025	0.237	0.77 (0.78, 1.22)	-2.1	0.88 (0.70, 1.30)	-0.8		
22 22	0.707	0.184	1.17 (0.78, 1.22)	1.5	1.06 (0.85, 1.15)	0.8		
23 23	0.951	0.190	1.06 (0.78, 1.22)	0.5	1.01 (0.83, 1.17)	0.1		
24 24	3.964	0.513	2.82 (0.78, 1.22)	10.9	1.09 (0.08, 1.92)	0.3		
25 25	2.142	0.252	2.97 (0.78, 1.22)	11.5	1.33 (0.67, 1.33)	2.0		
26 26	2.665	0.300	3.89 (0.78, 1.22)	15.1	1.25 (0.55, 1.45)	1.1		
27 27	-0.685	0.181	0.97 (0.78, 1.22)	-0.2	0.99 (0.86, 1.14)	-0.2		
28 28	4.265	0.589	6.99 (0.78, 1.22)	24.0	1.09 (0.00, 2.09)	0.3		
29 29	3.364	0.395	5.50 (0.78, 1.22)	20.1	1.16 (0.34, 1.66)	0.6		
30 30	2.968	0.337	3.10 (0.78, 1.22)	12.1	1.20 (0.47, 1.53)	0.8		
31 31	5.391	1.007	4.89 (0.78, 1.22)	18.3	1.05 (0.00, 2.95)	0.4		
32 32	3.964	0.513	4.60 (0.78, 1.22)	17.4	1.10 (0.08, 1.92)	0.4		
33 33	3.964	0.513	3.87 (0.78, 1.22)	15.0	1.10 (0.08, 1.92)	0.4		
34 34	3.364	0.395	4.90 (0.78, 1.22)	18.4	1.16 (0.34, 1.66)	0.6		
35 35	3.964	0.513	2.86 (0.78, 1.22)	11.0	1.05 (0.08, 1.92)	0.3		
36 36	2.759	0.311	2.96 (0.78, 1.22)	11.5	1.20 (0.53, 1.47)	0.9		
37 37	3.086	0.353	2.79 (0.78, 1.22)	10.7	1.18 (0.43, 1.57)	0.7		
38 38	4.684	0.717	6.26 (0.78, 1.22)	22.2	1.07 (0.00, 2.35)	0.3		
39 39	2.020	0.243	2.86 (0.78, 1.22)	11.0	1.36 (0.70, 1.30)	2.1		
40 40	2.496	0.283	3.47 (0.78, 1.22)	13.5	1.28 (0.60, 1.40)	1.3		
41 41	-3.783	0.462	0.35 (0.78, 1.22)	-7.7	0.91 (0.18, 1.82)	-0.1		
42 42	-1.669	0.215	0.91 (0.78, 1.22)	-0.7	0.95 (0.76, 1.24)	-0.4		
43 43	-3.273	0.372	0.46 (0.78, 1.22)	-5.9	0.91 (0.39, 1.61)	-0.2		
44 44	-2.203	0.250	0.82 (0.78, 1.22)	-1.7	0.88 (0.67, 1.33)	-0.7		
45 45	-1.716	0.218	0.73 (0.78, 1.22)	-2.5	0.90 (0.76, 1.24)	-0.7		
46 46	0.160	0.176	0.93 (0.78, 1.22)	-0.5	0.93 (0.88, 1.12)	-1.1		

47	47	-2.082	0.241	0.61 (0.78, 1.22)	-4.0	0.86 (0.69, 1.31)	-0.9
48	48	-2.816	0.310	0.65 (0.78, 1.22)	-3.4	0.92 (0.53, 1.47)	-0.3
49	49	-0.241	0.176	0.91 (0.78, 1.22)	-0.8	0.95 (0.88, 1.12)	-0.9
50	50	-3.783	0.462	0.36 (0.78, 1.22)	-7.6	0.91 (0.18, 1.82)	-0.1
51	51	-0.652	0.180	0.88 (0.78, 1.22)	-1.0	0.93 (0.87, 1.13)	-1.1
52	52	0.254	0.177	1.07 (0.78, 1.22)	0.6	1.05 (0.88, 1.12)	0.8
53	53	-1.098	0.191	0.71 (0.78, 1.22)	-2.8	0.83 (0.83, 1.17)	-2.1
54	54	-0.365	0.177	1.23 (0.78, 1.22)	1.9	1.18 (0.88, 1.12)	2.7
55	55	-0.272	0.176	0.78 (0.78, 1.22)	-2.1	0.83 (0.88, 1.12)	-2.9
56	56	-1.716	0.218	0.82 (0.78, 1.22)	-1.6	0.87 (0.76, 1.24)	-1.0
57	57	-0.428	0.177	0.74 (0.78, 1.22)	-2.4	0.78 (0.88, 1.12)	-3.7
58	58	-0.886	0.185	1.23 (0.78, 1.22)	1.9	1.12 (0.85, 1.15)	1.5
59	59	-2.477	0.274	0.55 (0.78, 1.22)	-4.7	0.88 (0.61, 1.39)	-0.6
60	60	0.476	0.180	0.87 (0.78, 1.22)	-1.1	0.88 (0.87, 1.13)	-1.8
61	61	-0.751	0.182	0.68 (0.78, 1.22)	-3.1	0.76 (0.86, 1.14)	-3.7
62	62	-0.751	0.182	1.15 (0.78, 1.22)	1.3	1.13 (0.86, 1.14)	1.8
63	63	-0.886	0.185	0.76 (0.78, 1.22)	-2.2	0.83 (0.85, 1.15)	-2.3
64	64	0.037	0.176	0.86 (0.78, 1.22)	-1.2	0.89 (0.88, 1.12)	-1.9
65	65	-0.334	0.177	0.99 (0.78, 1.22)	-0.1	1.01 (0.88, 1.12)	0.1
66	66	-1.172	0.194	1.13 (0.78, 1.22)	1.1	1.03 (0.82, 1.18)	0.3
67	67	-2.203	0.250	0.66 (0.78, 1.22)	-3.3	0.90 (0.67, 1.33)	-0.6
68	68	-1.026	0.189	0.77 (0.78, 1.22)	-2.2	0.87 (0.84, 1.16)	-1.6
69	69	-0.117	0.176	0.88 (0.78, 1.22)	-1.0	0.91 (0.88, 1.12)	-1.6
70	70	0.254	0.177	0.75 (0.78, 1.22)	-2.3	0.79 (0.88, 1.12)	-3.5
71	71	-1.172	0.194	0.74 (0.78, 1.22)	-2.5	0.86 (0.82, 1.18)	-1.6
72	72	-0.025	0.176	0.93 (0.78, 1.22)	-0.6	0.92 (0.88, 1.12)	-1.4
73	73	0.707	0.184	0.74 (0.78, 1.22)	-2.5	0.79 (0.85, 1.15)	-3.0
74	74	-0.491	0.178	0.92 (0.78, 1.22)	-0.7	0.93 (0.87, 1.13)	-1.1
75	75	0.574	0.181	0.93 (0.78, 1.22)	-0.6	0.92 (0.86, 1.14)	-1.1
76	76	-0.555	0.179	0.91 (0.78, 1.22)	-0.7	0.95 (0.87, 1.13)	-0.8
77	77	-0.056	0.176	1.09 (0.78, 1.22)	0.8	1.08 (0.88, 1.12)	1.3
78	78	0.640	0.182	0.91 (0.78, 1.22)	-0.8	0.92 (0.86, 1.14)	-1.1
79	79	-2.477	0.274	0.57 (0.78, 1.22)	-4.5	0.86 (0.61, 1.39)	-0.7
80	80	-0.990	0.188	1.00 (0.78, 1.22)	0.1	1.02 (0.84, 1.16)	0.2
81	81	0.068	0.176	1.03 (0.78, 1.22)	0.3	1.04 (0.88, 1.12)	0.7
82	82	0.809	0.186	1.03 (0.78, 1.22)	0.3	0.98 (0.85, 1.15)	-0.2
83	83	-0.087	0.176	1.00 (0.78, 1.22)	0.0	1.01 (0.88, 1.12)	0.2
84	84	-0.491	0.178	0.99 (0.78, 1.22)	-0.1	1.02 (0.87, 1.13)	0.3
85	85	-1.578	0.211	1.09 (0.78, 1.22)	0.8	1.06 (0.77, 1.23)	0.5
86	86	-1.026	0.189	1.01 (0.78, 1.22)	0.1	1.05 (0.84, 1.16)	0.6
87	87	-0.025	0.176	0.96 (0.78, 1.22)	-0.3	0.98 (0.88, 1.12)	-0.3
88	88	-0.491	0.178	1.03 (0.78, 1.22)	0.3	1.06 (0.87, 1.13)	0.9
89	89	2.142	0.252	1.70 (0.78, 1.22)	5.1	1.24 (0.67, 1.33)	1.4
90	90	0.254	0.177	1.31 (0.78, 1.22)	2.5	1.25 (0.88, 1.12)	3.6
91	91	-0.955	0.187	0.91 (0.78, 1.22)	-0.8	0.96 (0.84, 1.16)	-0.4
92	92	-0.555	0.179	0.87 (0.78, 1.22)	-1.2	0.91 (0.87, 1.13)	-1.3
93	93	0.741	0.185	1.00 (0.78, 1.22)	0.0	0.96 (0.85, 1.15)	-0.5
94	94	0.285	0.177	0.88 (0.78, 1.22)	-1.1	0.90 (0.87, 1.13)	-1.6
95	95	0.380	0.178	1.29 (0.78, 1.22)	2.4	1.19 (0.87, 1.13)	2.7
96	96	1.602	0.217	1.09 (0.78, 1.22)	0.8	0.99 (0.76, 1.24)	-0.1
97	97	-0.117	0.176	0.97 (0.78, 1.22)	-0.2	0.99 (0.88, 1.12)	-0.2
98	98	-0.460	0.178	1.07 (0.78, 1.22)	0.6	1.02 (0.87, 1.13)	0.4
99	99	-1.026	0.189	0.94 (0.78, 1.22)	-0.5	0.98 (0.84, 1.16)	-0.2
100	100	1.024	0.192	1.28 (0.78, 1.22)	2.3	1.09 (0.83, 1.17)	1.0

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.978

Chi-square test of parameter equality = 3685.17, df = 100, Sig Level = 0.000

^ Quick standard errors have been used

Kết quả phân tích sự phù hợp với mô hình của Đề 2

ConQuest: Generalised Item Response Modelling Software Fri Feb 03 21:03 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1 1	-2.575	0.290	1.21 (0.77, 1.23)	1.8	1.01 (0.58, 1.42)	0.1		
2 2	-3.525	0.424	0.76 (0.77, 1.23)	-2.2	0.88 (0.29, 1.71)	-0.2		
3 3	-5.382	1.007	1.73 (0.77, 1.23)	5.3	1.00 (0.00, 2.91)	0.3		
4 4	-3.720	0.461	0.81 (0.77, 1.23)	-1.7	0.93 (0.21, 1.79)	-0.0		
5 5	-4.257	0.589	0.24 (0.77, 1.23)	-9.8	0.89 (0.00, 2.06)	-0.0		
6 6	0.048	0.176	1.19 (0.77, 1.23)	1.6	1.16 (0.88, 1.12)	2.5		
7 7	-3.525	0.424	0.56 (0.77, 1.23)	-4.6	0.91 (0.29, 1.71)	-0.1		
8 8	-1.105	0.194	1.08 (0.77, 1.23)	0.7	1.06 (0.82, 1.18)	0.6		
9 9	-1.384	0.205	1.31 (0.77, 1.23)	2.5	1.21 (0.79, 1.21)	1.9		
10 10	-2.662	0.300	1.11 (0.77, 1.23)	1.0	0.97 (0.56, 1.44)	-0.1		
11 11	-1.749	0.224	0.85 (0.77, 1.23)	-1.3	0.95 (0.74, 1.26)	-0.4		
13 13	-0.263	0.177	1.14 (0.77, 1.23)	1.2	1.12 (0.88, 1.12)	1.8		
14 14	0.361	0.178	1.41 (0.77, 1.23)	3.2	1.24 (0.87, 1.13)	3.5		
15 15	-0.200	0.177	1.08 (0.77, 1.23)	0.7	1.07 (0.88, 1.12)	1.1		
16 16	-3.956	0.513	0.55 (0.77, 1.23)	-4.6	0.94 (0.10, 1.90)	0.0		
17 17	-0.326	0.178	1.19 (0.77, 1.23)	1.6	1.14 (0.88, 1.12)	2.1		
18 18	-0.453	0.179	0.83 (0.77, 1.23)	-1.5	0.88 (0.87, 1.13)	-1.8		
19 19	-3.956	0.513	0.59 (0.77, 1.23)	-4.2	0.98 (0.10, 1.90)	0.1		
20 20	-2.575	0.290	0.93 (0.77, 1.23)	-0.6	1.00 (0.58, 1.42)	0.1		
21 21	-2.205	0.256	1.35 (0.77, 1.23)	2.8	1.04 (0.66, 1.34)	0.3		
22 22	-0.715	0.183	0.87 (0.77, 1.23)	-1.1	0.94 (0.86, 1.14)	-0.9		
23 23	-2.963	0.336	0.46 (0.77, 1.23)	-6.0	0.86 (0.48, 1.52)	-0.5		
24 24	0.079	0.176	1.19 (0.77, 1.23)	1.6	1.14 (0.88, 1.12)	2.2		
25 25	-1.384	0.205	1.21 (0.77, 1.23)	1.8	1.05 (0.79, 1.21)	0.5		
26 26	-2.755	0.310	1.01 (0.77, 1.23)	0.1	0.94 (0.54, 1.46)	-0.2		
27 27	-0.232	0.177	0.90 (0.77, 1.23)	-0.9	0.91 (0.88, 1.12)	-1.5		
28 28	-2.020	0.242	0.54 (0.77, 1.23)	-4.8	0.77 (0.70, 1.30)	-1.6		
29 29	-0.582	0.181	0.92 (0.77, 1.23)	-0.7	0.96 (0.86, 1.14)	-0.6		
30 30	-1.513	0.211	0.97 (0.77, 1.23)	-0.3	0.98 (0.78, 1.22)	-0.1		
31 31	-1.605	0.216	0.80 (0.77, 1.23)	-1.8	0.89 (0.76, 1.24)	-0.9		
32 32	-0.615	0.182	0.98 (0.77, 1.23)	-0.1	1.01 (0.86, 1.14)	0.1		
33 33	0.361	0.178	1.00 (0.77, 1.23)	0.1	0.95 (0.87, 1.13)	-0.7		
34 34	-1.749	0.224	0.87 (0.77, 1.23)	-1.2	0.93 (0.74, 1.26)	-0.5		
35 35	0.235	0.177	1.03 (0.77, 1.23)	0.3	1.01 (0.88, 1.12)	0.2		
36 36	-0.389	0.178	0.85 (0.77, 1.23)	-1.3	0.88 (0.87, 1.13)	-1.9		
37 37	-3.212	0.371	0.32 (0.77, 1.23)	-8.2	0.82 (0.40, 1.60)	-0.5		
38 38	-0.357	0.178	0.86 (0.77, 1.23)	-1.2	0.87 (0.87, 1.13)	-2.1		
39 39	-0.994	0.191	0.94 (0.77, 1.23)	-0.5	0.94 (0.83, 1.17)	-0.7		
40 40	-1.342	0.203	0.77 (0.77, 1.23)	-2.1	0.87 (0.80, 1.20)	-1.2		
41 41	-3.358	0.395	0.57 (0.77, 1.23)	-4.4	0.86 (0.35, 1.65)	-0.3		
42 42	-3.720	0.461	0.32 (0.77, 1.23)	-8.2	0.89 (0.21, 1.79)	-0.2		
43 43	-3.212	0.371	1.17 (0.77, 1.23)	1.5	0.90 (0.40, 1.60)	-0.3		
44 44	0.329	0.178	1.17 (0.77, 1.23)	1.5	1.11 (0.88, 1.12)	1.7		
45 45	-0.648	0.182	1.01 (0.77, 1.23)	0.1	1.03 (0.86, 1.14)	0.4		
46 46	-0.453	0.179	0.82 (0.77, 1.23)	-1.6	0.86 (0.87, 1.13)	-2.2		
47 47	-1.652	0.218	0.78 (0.77, 1.23)	-2.0	0.91 (0.76, 1.24)	-0.7		
48 48	-2.079	0.246	0.72 (0.77, 1.23)	-2.6	0.88 (0.69, 1.31)	-0.8		
49 49	0.298	0.177	0.96 (0.77, 1.23)	-0.3	0.97 (0.88, 1.12)	-0.5		

50	50	-2.493	0.282	0.56 (0.77, 1.23)	-4.5	0.87 (0.60, 1.40)	-0.6
51	51	0.079	0.176	0.96 (0.77, 1.23)	-0.3	0.96 (0.88, 1.12)	-0.6
52	52	-0.014	0.176	1.43 (0.77, 1.23)	3.4	1.33 (0.88, 1.12)	5.0
53	53	-0.582	0.181	0.82 (0.77, 1.23)	-1.6	0.87 (0.86, 1.14)	-1.9
54	54	-2.575	0.290	1.10 (0.77, 1.23)	0.8	0.90 (0.58, 1.42)	-0.4
55	55	-0.045	0.176	0.96 (0.77, 1.23)	-0.3	0.97 (0.88, 1.12)	-0.5
56	56	-0.326	0.178	1.04 (0.77, 1.23)	0.4	1.03 (0.88, 1.12)	0.4
57	57	0.017	0.176	0.91 (0.77, 1.23)	-0.8	0.92 (0.88, 1.12)	-1.4
58	58	0.079	0.176	0.89 (0.77, 1.23)	-0.9	0.90 (0.88, 1.12)	-1.7
59	59	-1.652	0.218	0.64 (0.77, 1.23)	-3.6	0.82 (0.76, 1.24)	-1.6
60	60	-1.513	0.211	0.90 (0.77, 1.23)	-0.9	0.88 (0.78, 1.22)	-1.1
61	61	-0.648	0.182	0.77 (0.77, 1.23)	-2.1	0.83 (0.86, 1.14)	-2.5
62	62	-1.907	0.234	0.55 (0.77, 1.23)	-4.7	0.80 (0.72, 1.28)	-1.5
63	63	-0.749	0.184	0.84 (0.77, 1.23)	-1.5	0.85 (0.85, 1.15)	-2.1
64	64	0.110	0.176	1.11 (0.77, 1.23)	1.0	1.10 (0.88, 1.12)	1.6
65	65	-1.301	0.202	0.86 (0.77, 1.23)	-1.3	0.91 (0.80, 1.20)	-0.9
66	66	-1.031	0.192	1.18 (0.77, 1.23)	1.5	1.03 (0.83, 1.17)	0.3
67	67	-1.800	0.227	0.70 (0.77, 1.23)	-2.9	0.85 (0.73, 1.27)	-1.1
68	68	-2.079	0.246	0.77 (0.77, 1.23)	-2.1	0.91 (0.69, 1.31)	-0.5
69	69	0.652	0.182	1.12 (0.77, 1.23)	1.0	0.99 (0.86, 1.14)	-0.1
70	70	0.266	0.177	0.96 (0.77, 1.23)	-0.3	0.95 (0.88, 1.12)	-0.8
71	71	1.036	0.192	1.32 (0.77, 1.23)	2.6	1.05 (0.83, 1.17)	0.6
72	72	1.565	0.214	1.83 (0.77, 1.23)	5.9	1.19 (0.77, 1.23)	1.5
73	73	0.927	0.189	1.36 (0.77, 1.23)	2.8	1.15 (0.84, 1.16)	1.8
74	74	0.017	0.176	0.85 (0.77, 1.23)	-1.3	0.89 (0.88, 1.12)	-1.8
75	75	0.891	0.188	1.27 (0.77, 1.23)	2.2	1.16 (0.84, 1.16)	1.9
76	76	0.298	0.177	1.11 (0.77, 1.23)	1.0	1.07 (0.88, 1.12)	1.1
77	77	-0.852	0.187	1.07 (0.77, 1.23)	0.7	1.03 (0.84, 1.16)	0.4
78	78	0.424	0.179	1.23 (0.77, 1.23)	1.9	1.09 (0.87, 1.13)	1.3
79	79	-1.384	0.205	0.89 (0.77, 1.23)	-1.0	0.89 (0.79, 1.21)	-1.1
80	80	-0.045	0.176	0.92 (0.77, 1.23)	-0.7	0.95 (0.88, 1.12)	-0.8
81	81	-0.852	0.187	0.95 (0.77, 1.23)	-0.4	1.01 (0.84, 1.16)	0.1
82	82	-0.107	0.176	1.15 (0.77, 1.23)	1.3	1.08 (0.88, 1.12)	1.2
83	83	0.999	0.191	1.13 (0.77, 1.23)	1.1	1.08 (0.83, 1.17)	0.9
84	84	0.204	0.177	1.15 (0.77, 1.23)	1.3	1.13 (0.88, 1.12)	2.0
85	85	-0.138	0.176	0.97 (0.77, 1.23)	-0.3	0.96 (0.88, 1.12)	-0.6
86	86	-0.357	0.178	1.21 (0.77, 1.23)	1.7	1.11 (0.87, 1.13)	1.7
87	87	0.392	0.178	1.30 (0.77, 1.23)	2.4	1.20 (0.87, 1.13)	2.9
88	88	-0.263	0.177	1.06 (0.77, 1.23)	0.5	1.06 (0.88, 1.12)	0.9
89	89	0.298	0.177	1.05 (0.77, 1.23)	0.5	1.01 (0.88, 1.12)	0.2
90	90	1.390	0.206	0.85 (0.77, 1.23)	-1.3	0.93 (0.79, 1.21)	-0.7
91	91	0.266	0.177	1.04 (0.77, 1.23)	0.4	1.05 (0.88, 1.12)	0.9
92	92	-0.200	0.177	1.01 (0.77, 1.23)	0.2	1.01 (0.88, 1.12)	0.1
93	93	0.521	0.180	1.16 (0.77, 1.23)	1.4	1.06 (0.87, 1.13)	0.8
94	94	0.488	0.180	1.08 (0.77, 1.23)	0.7	1.05 (0.87, 1.13)	0.7
95	95	1.149	0.196	1.40 (0.77, 1.23)	3.2	1.18 (0.82, 1.18)	1.8
96	96	1.036	0.192	1.06 (0.77, 1.23)	0.5	1.02 (0.83, 1.17)	0.3
97	97	0.048	0.176	1.12 (0.77, 1.23)	1.0	1.07 (0.88, 1.12)	1.2
98	98	-0.232	0.177	0.99 (0.77, 1.23)	-0.0	0.99 (0.88, 1.12)	-0.1
99	99	-0.200	0.177	1.05 (0.77, 1.23)	0.5	1.02 (0.88, 1.12)	0.3
100	100	1.612	0.216	1.28 (0.77, 1.23)	2.3	0.97 (0.76, 1.24)	-0.2

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.968

Chi-square test of parameter equality = 2925.76, df = 99, Sig Level = 0.000

^ Quick standard errors have been used

Kết quả phân tích sự phù hợp với mô hình của Đề 3

ConQuest: Generalised Item Response Modelling Software Sun Feb 05 10:24 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1 1	-5.623	1.008	2.01 (0.78, 1.22)	7.0	1.00 (0.00, 2.89)	0.3		
2 2	-2.088	0.240	2.63 (0.78, 1.22)	10.1	1.27 (0.72, 1.28)	1.8		
3 3	-2.207	0.248	1.01 (0.78, 1.22)	0.1	1.05 (0.71, 1.29)	0.4		
4 4	-3.756	0.427	1.45 (0.78, 1.22)	3.5	0.96 (0.30, 1.70)	-0.0		
5 5	-5.623	1.008	0.25 (0.78, 1.22)	-9.7	0.96 (0.00, 2.89)	0.3		
6 6	-1.976	0.234	2.43 (0.78, 1.22)	9.1	1.30 (0.74, 1.26)	2.1		
7 7	-1.719	0.221	1.30 (0.78, 1.22)	2.5	1.07 (0.77, 1.23)	0.6		
8 8	-5.623	1.008	1.73 (0.78, 1.22)	5.3	1.00 (0.00, 2.89)	0.3		
9 9	-3.953	0.464	1.20 (0.78, 1.22)	1.7	1.01 (0.22, 1.78)	0.2		
10 10	-0.730	0.191	1.09 (0.78, 1.22)	0.8	1.03 (0.83, 1.17)	0.4		
11 11	0.178	0.185	1.70 (0.78, 1.22)	5.2	1.46 (0.84, 1.16)	5.0		
12 12	-4.494	0.590	0.71 (0.78, 1.22)	-2.8	0.94 (0.00, 2.05)	0.1		
13 13	-4.915	0.718	1.51 (0.78, 1.22)	3.9	0.98 (0.00, 2.31)	0.2		
14 14	0.109	0.185	1.19 (0.78, 1.22)	1.6	1.03 (0.84, 1.16)	0.4		
15 15	-1.719	0.221	0.82 (0.78, 1.22)	-1.7	0.89 (0.77, 1.23)	-1.0		
16 16	-3.756	0.427	0.34 (0.78, 1.22)	-7.9	0.90 (0.30, 1.70)	-0.2		
17 17	-1.818	0.225	1.19 (0.78, 1.22)	1.6	1.01 (0.76, 1.24)	0.1		
18 18	-0.730	0.191	1.34 (0.78, 1.22)	2.8	1.14 (0.83, 1.17)	1.6		
19 19	-2.207	0.248	0.71 (0.78, 1.22)	-2.8	0.90 (0.71, 1.29)	-0.6		
20 20	-3.304	0.356	0.64 (0.78, 1.22)	-3.6	0.95 (0.46, 1.54)	-0.1		
21 21	-3.304	0.356	0.59 (0.78, 1.22)	-4.2	0.94 (0.46, 1.54)	-0.1		
22 22	-0.916	0.195	1.44 (0.78, 1.22)	3.4	1.21 (0.83, 1.17)	2.2		
23 23	-1.768	0.223	1.42 (0.78, 1.22)	3.3	1.07 (0.76, 1.24)	0.6		
24 24	-2.088	0.240	0.93 (0.78, 1.22)	-0.6	0.92 (0.72, 1.28)	-0.6		
25 25	-2.031	0.237	0.51 (0.78, 1.22)	-5.3	0.83 (0.73, 1.27)	-1.3		
26 26	0.006	0.185	1.33 (0.78, 1.22)	2.7	1.25 (0.84, 1.16)	2.9		
27 27	-2.700	0.287	0.97 (0.78, 1.22)	-0.3	0.93 (0.62, 1.38)	-0.3		
28 28	-1.577	0.215	0.92 (0.78, 1.22)	-0.7	1.01 (0.78, 1.22)	0.1		
29 29	-0.063	0.185	0.99 (0.78, 1.22)	-0.0	0.96 (0.84, 1.16)	-0.5		
30 30	-0.409	0.187	1.02 (0.78, 1.22)	0.2	1.02 (0.84, 1.16)	0.3		
31 31	-1.818	0.225	0.56 (0.78, 1.22)	-4.6	0.81 (0.76, 1.24)	-1.6		
32 32	-0.804	0.192	1.29 (0.78, 1.22)	2.4	1.18 (0.83, 1.17)	2.1		
33 33	-2.785	0.295	0.60 (0.78, 1.22)	-4.1	0.89 (0.60, 1.40)	-0.5		
34 34	0.632	0.189	1.32 (0.78, 1.22)	2.6	1.13 (0.84, 1.16)	1.5		
35 35	-1.272	0.204	0.68 (0.78, 1.22)	-3.1	0.86 (0.81, 1.19)	-1.5		
36 36	-3.183	0.340	0.35 (0.78, 1.22)	-7.8	0.82 (0.50, 1.50)	-0.7		
37 37	-1.314	0.205	0.71 (0.78, 1.22)	-2.8	0.83 (0.81, 1.19)	-1.7		
38 38	-2.401	0.262	1.05 (0.78, 1.22)	0.5	0.81 (0.67, 1.33)	-1.1		
39 39	-1.486	0.211	0.91 (0.78, 1.22)	-0.8	0.94 (0.79, 1.21)	-0.6		
40 40	-1.531	0.213	1.26 (0.78, 1.22)	2.1	1.13 (0.79, 1.21)	1.2		
41 41	-4.192	0.515	0.53 (0.78, 1.22)	-5.1	0.96 (0.11, 1.89)	0.1		
42 42	0.178	0.185	0.93 (0.78, 1.22)	-0.6	0.94 (0.84, 1.16)	-0.7		
43 43	-0.270	0.186	1.03 (0.78, 1.22)	0.3	1.05 (0.84, 1.16)	0.6		
44 44	-1.577	0.215	0.55 (0.78, 1.22)	-4.8	0.75 (0.78, 1.22)	-2.4		
45 45	-0.993	0.196	0.60 (0.78, 1.22)	-4.2	0.75 (0.82, 1.18)	-3.0		
46 46	-2.700	0.287	1.05 (0.78, 1.22)	0.5	0.98 (0.62, 1.38)	-0.0		
47 47	-2.088	0.240	0.73 (0.78, 1.22)	-2.6	0.99 (0.72, 1.28)	0.0		
48 48	1.193	0.199	0.91 (0.78, 1.22)	-0.8	0.95 (0.82, 1.18)	-0.6		

49 49	-2.620	0.280	1.29 (0.78, 1.22)	2.4	0.98 (0.63, 1.37)	-0.1
50 50	-2.471	0.267	1.21 (0.78, 1.22)	1.8	1.02 (0.66, 1.34)	0.2
51 51	-0.804	0.192	0.89 (0.78, 1.22)	-0.9	0.96 (0.83, 1.17)	-0.4
52 52	-1.719	0.221	0.59 (0.78, 1.22)	-4.2	0.85 (0.77, 1.23)	-1.3
53 53	-3.586	0.398	1.39 (0.78, 1.22)	3.1	0.99 (0.37, 1.63)	0.1
54 54	-2.031	0.237	0.53 (0.78, 1.22)	-5.0	0.81 (0.73, 1.27)	-1.4
55 55	-0.622	0.190	0.91 (0.78, 1.22)	-0.8	0.93 (0.84, 1.16)	-0.8
56 56	-0.767	0.192	0.95 (0.78, 1.22)	-0.4	1.05 (0.83, 1.17)	0.6
57 57	-0.841	0.193	0.72 (0.78, 1.22)	-2.7	0.86 (0.83, 1.17)	-1.7
58 58	-0.586	0.189	0.93 (0.78, 1.22)	-0.6	1.02 (0.84, 1.16)	0.3
59 59	-2.334	0.257	0.86 (0.78, 1.22)	-1.2	0.96 (0.69, 1.31)	-0.2
60 60	-3.304	0.356	0.58 (0.78, 1.22)	-4.4	0.88 (0.46, 1.54)	-0.4
61 61	-2.270	0.252	0.48 (0.78, 1.22)	-5.7	0.80 (0.70, 1.30)	-1.4
62 62	-0.339	0.187	0.88 (0.78, 1.22)	-1.1	0.87 (0.84, 1.16)	-1.7
63 63	-0.767	0.192	1.02 (0.78, 1.22)	0.2	1.03 (0.83, 1.17)	0.4
64 64	-0.166	0.186	1.10 (0.78, 1.22)	0.9	1.05 (0.84, 1.16)	0.6
65 65	-0.767	0.192	0.73 (0.78, 1.22)	-2.6	0.84 (0.83, 1.17)	-1.9
66 66	-0.841	0.193	0.71 (0.78, 1.22)	-2.8	0.76 (0.83, 1.17)	-3.1
67 67	-1.356	0.206	0.87 (0.78, 1.22)	-1.2	0.96 (0.80, 1.20)	-0.3
68 68	-1.486	0.211	0.99 (0.78, 1.22)	-0.1	0.98 (0.79, 1.21)	-0.2
69 69	-0.374	0.187	1.18 (0.78, 1.22)	1.6	1.17 (0.84, 1.16)	2.0
70 70	0.596	0.189	0.90 (0.78, 1.22)	-0.9	0.83 (0.84, 1.16)	-2.2
71 71	-2.970	0.314	1.16 (0.78, 1.22)	1.4	0.89 (0.55, 1.45)	-0.4
72 72	-1.442	0.209	0.88 (0.78, 1.22)	-1.1	1.04 (0.80, 1.20)	0.4
73 73	-0.804	0.192	0.68 (0.78, 1.22)	-3.1	0.80 (0.83, 1.17)	-2.4
74 74	-0.878	0.194	1.20 (0.78, 1.22)	1.7	1.15 (0.83, 1.17)	1.6
75 75	-0.515	0.188	1.09 (0.78, 1.22)	0.8	1.11 (0.84, 1.16)	1.3
76 76	-1.486	0.211	0.81 (0.78, 1.22)	-1.7	0.99 (0.79, 1.21)	-0.1
77 77	0.143	0.185	1.24 (0.78, 1.22)	2.0	1.11 (0.84, 1.16)	1.3
78 78	-2.334	0.257	0.90 (0.78, 1.22)	-0.9	0.90 (0.69, 1.31)	-0.6
79 79	-1.150	0.200	1.19 (0.78, 1.22)	1.6	1.18 (0.82, 1.18)	1.9
80 80	-2.401	0.262	0.51 (0.78, 1.22)	-5.2	0.88 (0.67, 1.33)	-0.7
81 81	-1.486	0.211	0.78 (0.78, 1.22)	-2.0	0.84 (0.79, 1.21)	-1.5
82 82	-0.954	0.195	1.11 (0.78, 1.22)	1.0	1.03 (0.83, 1.17)	0.3
83 83	-1.976	0.234	0.75 (0.78, 1.22)	-2.4	0.98 (0.74, 1.26)	-0.1
84 84	-0.993	0.196	0.99 (0.78, 1.22)	-0.0	1.01 (0.82, 1.18)	0.2
85 85	-0.339	0.187	0.71 (0.78, 1.22)	-2.8	0.80 (0.84, 1.16)	-2.6
86 86	-0.270	0.186	0.97 (0.78, 1.22)	-0.2	0.99 (0.84, 1.16)	-0.1
87 87	-1.623	0.216	0.76 (0.78, 1.22)	-2.3	0.94 (0.78, 1.22)	-0.5
88 88	-1.314	0.205	0.80 (0.78, 1.22)	-1.9	0.95 (0.81, 1.19)	-0.5
89 89	2.112	0.235	4.38 (0.78, 1.22)	16.9	1.65 (0.74, 1.26)	4.1
90 90	-0.550	0.189	0.75 (0.78, 1.22)	-2.3	0.85 (0.84, 1.16)	-1.9
91 91	-0.586	0.189	0.86 (0.78, 1.22)	-1.3	0.92 (0.84, 1.16)	-1.0
92 92	-1.314	0.205	0.70 (0.78, 1.22)	-2.9	0.88 (0.81, 1.19)	-1.2
93 93	-0.028	0.185	1.15 (0.78, 1.22)	1.3	1.11 (0.84, 1.16)	1.3
94 94	0.247	0.186	0.99 (0.78, 1.22)	-0.0	0.99 (0.84, 1.16)	-0.2
95 95	-1.031	0.197	1.17 (0.78, 1.22)	1.5	1.04 (0.82, 1.18)	0.5
96 96	2.932	0.297	4.91 (0.78, 1.22)	18.5	1.34 (0.59, 1.41)	1.5
97 97	-1.110	0.199	0.89 (0.78, 1.22)	-1.0	0.97 (0.82, 1.18)	-0.3
98 98	-0.916	0.195	1.10 (0.78, 1.22)	0.9	1.09 (0.83, 1.17)	1.1
99 99	-1.314	0.205	0.73 (0.78, 1.22)	-2.6	0.89 (0.81, 1.19)	-1.1
100 100	-0.694	0.191	0.89 (0.78, 1.22)	-1.0	0.97 (0.84, 1.16)	-0.3

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.958

Chi-square test of parameter equality = 4342.52, df = 100, Sig Level = 0.000

^ Quick standard errors have been used

Kết quả phân tích sự phù hợp với mô hình của Đề 4

ConQuest: Generalised Item Response Modelling Software Tue Feb 14 11:48 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
2 2	-3.069	0.395	2.85 (0.77, 1.23)	10.9	1.06 (0.34, 1.66)	0.3		
3 3	-3.069	0.395	1.20 (0.77, 1.23)	1.6	1.03 (0.34, 1.66)	0.2		
4 4	-3.236	0.424	3.02 (0.77, 1.23)	11.6	1.08 (0.27, 1.73)	0.3		
5 5	-3.968	0.589	0.59 (0.77, 1.23)	-4.1	1.00 (0.00, 2.08)	0.2		
6 6	-1.205	0.214	0.83 (0.77, 1.23)	-1.5	0.92 (0.79, 1.21)	-0.7		
7 7	-5.093	1.006	0.38 (0.77, 1.23)	-7.1	1.00 (0.00, 2.94)	0.3		
8 8	-3.968	0.589	0.90 (0.77, 1.23)	-0.9	1.01 (0.00, 2.08)	0.2		
9 9	-0.637	0.195	0.81 (0.77, 1.23)	-1.7	0.93 (0.84, 1.16)	-0.9		
10 10	-3.199	0.423	1.60 (0.77, 1.23)	4.4	1.00 (0.28, 1.72)	0.1		
11 11	-1.447	0.226	1.07 (0.77, 1.23)	0.7	0.99 (0.75, 1.25)	-0.0		
12 12	-3.431	0.462	0.47 (0.77, 1.23)	-5.8	0.96 (0.19, 1.81)	0.0		
13 13	-3.236	0.424	1.17 (0.77, 1.23)	1.4	1.01 (0.27, 1.73)	0.2		
14 14	-1.910	0.257	1.22 (0.77, 1.23)	1.8	1.13 (0.67, 1.33)	0.8		
15 15	-4.387	0.716	0.30 (0.77, 1.23)	-8.5	0.96 (0.00, 2.35)	0.2		
16 16	-2.673	0.336	0.59 (0.77, 1.23)	-4.1	0.96 (0.47, 1.53)	-0.1		
17 17	-0.591	0.193	0.77 (0.77, 1.23)	-2.1	0.90 (0.85, 1.15)	-1.4		
18 18	-0.091	0.186	1.44 (0.77, 1.23)	3.4	1.23 (0.87, 1.13)	3.2		
19 19	0.149	0.185	1.21 (0.77, 1.23)	1.8	1.16 (0.87, 1.13)	2.3		
20 20	0.047	0.185	1.14 (0.77, 1.23)	1.2	1.11 (0.87, 1.13)	1.6		
21 21	-5.093	1.006	0.25 (0.77, 1.23)	-9.6	0.98 (0.00, 2.94)	0.3		
22 22	-0.704	0.196	1.13 (0.77, 1.23)	1.1	1.05 (0.84, 1.16)	0.7		
23 23	1.027	0.193	1.19 (0.77, 1.23)	1.6	1.11 (0.83, 1.17)	1.3		
24 24	-1.299	0.219	0.98 (0.77, 1.23)	-0.2	0.96 (0.77, 1.23)	-0.3		
25 25	2.011	0.226	1.19 (0.77, 1.23)	1.6	1.01 (0.75, 1.25)	0.1		
26 26	-1.977	0.263	0.65 (0.77, 1.23)	-3.4	0.90 (0.65, 1.35)	-0.5		
27 27	0.286	0.185	1.16 (0.77, 1.23)	1.3	1.04 (0.86, 1.14)	0.7		
28 28	0.561	0.186	0.90 (0.77, 1.23)	-0.9	0.92 (0.86, 1.14)	-1.1		
29 29	-0.480	0.191	0.82 (0.77, 1.23)	-1.7	0.94 (0.85, 1.15)	-0.8		
30 30	0.772	0.189	1.02 (0.77, 1.23)	0.2	1.02 (0.84, 1.16)	0.2		
31 31	0.880	0.191	1.03 (0.77, 1.23)	0.3	0.99 (0.84, 1.16)	-0.1		
32 32	1.027	0.193	1.09 (0.77, 1.23)	0.8	1.07 (0.83, 1.17)	0.8		
33 33	-2.370	0.300	2.13 (0.77, 1.23)	7.5	1.05 (0.56, 1.44)	0.3		
34 34	0.081	0.185	0.97 (0.77, 1.23)	-0.3	1.02 (0.87, 1.13)	0.3		
35 35	-3.431	0.462	0.32 (0.77, 1.23)	-8.2	0.90 (0.19, 1.81)	-0.1		
36 36	-0.300	0.188	1.07 (0.77, 1.23)	0.6	1.09 (0.86, 1.14)	1.3		
37 37	-1.782	0.248	0.55 (0.77, 1.23)	-4.7	0.83 (0.69, 1.31)	-1.1		
38 38	0.457	0.186	1.02 (0.77, 1.23)	0.2	1.02 (0.86, 1.14)	0.2		
39 39	-0.944	0.204	1.01 (0.77, 1.23)	0.1	1.02 (0.82, 1.18)	0.3		
40 40	-1.028	0.207	0.80 (0.77, 1.23)	-1.8	0.94 (0.81, 1.19)	-0.6		
41 41	0.916	0.191	1.02 (0.77, 1.23)	0.2	1.05 (0.84, 1.16)	0.7		
42 42	-3.431	0.462	0.39 (0.77, 1.23)	-7.0	0.93 (0.19, 1.81)	-0.0		
43 43	-0.986	0.205	0.65 (0.77, 1.23)	-3.4	0.81 (0.81, 1.19)	-2.0		
44 44	-0.230	0.187	0.76 (0.77, 1.23)	-2.3	0.85 (0.87, 1.13)	-2.2		
45 45	-2.464	0.311	1.19 (0.77, 1.23)	1.6	0.93 (0.54, 1.46)	-0.2		
46 46	-2.673	0.336	0.85 (0.77, 1.23)	-1.4	0.99 (0.47, 1.53)	0.0		
47 47	1.179	0.197	0.96 (0.77, 1.23)	-0.4	0.93 (0.82, 1.18)	-0.7		
48 48	-0.516	0.192	0.89 (0.77, 1.23)	-1.0	0.98 (0.85, 1.15)	-0.3		
49 49	-1.299	0.219	1.05 (0.77, 1.23)	0.4	1.04 (0.77, 1.23)	0.3		

50	50	0.843	0.190	1.11 (0.77, 1.23)	0.9	1.02 (0.84, 1.16)	0.3
51	51	-1.205	0.214	0.85 (0.77, 1.23)	-1.4	0.99 (0.79, 1.21)	-0.1
52	52	1.218	0.198	1.28 (0.77, 1.23)	2.3	1.09 (0.81, 1.19)	1.0
53	53	-1.664	0.240	0.69 (0.77, 1.23)	-3.0	0.89 (0.72, 1.28)	-0.8
54	54	-0.407	0.190	0.82 (0.77, 1.23)	-1.7	0.94 (0.86, 1.14)	-0.9
55	55	-0.230	0.187	0.98 (0.77, 1.23)	-0.1	1.05 (0.87, 1.13)	0.7
56	56	-0.056	0.186	0.92 (0.77, 1.23)	-0.6	0.97 (0.87, 1.13)	-0.5
57	57	-0.125	0.186	0.90 (0.77, 1.23)	-0.8	0.97 (0.87, 1.13)	-0.4
58	58	-1.347	0.221	0.58 (0.77, 1.23)	-4.3	0.79 (0.77, 1.23)	-1.8
59	59	-2.122	0.276	0.95 (0.77, 1.23)	-0.4	0.86 (0.62, 1.38)	-0.7
60	60	-0.371	0.189	0.76 (0.77, 1.23)	-2.2	0.88 (0.86, 1.14)	-1.7
61	61	0.047	0.185	0.80 (0.77, 1.23)	-1.8	0.87 (0.87, 1.13)	-1.9
62	62	0.491	0.186	0.95 (0.77, 1.23)	-0.4	0.96 (0.86, 1.14)	-0.5
63	63	0.286	0.185	0.91 (0.77, 1.23)	-0.7	0.96 (0.86, 1.14)	-0.5
64	64	0.320	0.185	1.00 (0.77, 1.23)	0.0	1.03 (0.86, 1.14)	0.4
65	65	0.081	0.185	0.76 (0.77, 1.23)	-2.2	0.85 (0.87, 1.13)	-2.4
66	66	0.081	0.185	0.97 (0.77, 1.23)	-0.2	1.01 (0.87, 1.13)	0.2
67	67	0.149	0.185	0.87 (0.77, 1.23)	-1.1	0.93 (0.87, 1.13)	-1.0
68	68	0.423	0.185	1.05 (0.77, 1.23)	0.5	1.05 (0.86, 1.14)	0.8
69	69	-1.071	0.208	0.98 (0.77, 1.23)	-0.1	1.01 (0.80, 1.20)	0.1
70	70	-1.028	0.207	0.75 (0.77, 1.23)	-2.3	0.88 (0.81, 1.19)	-1.3
71	71	-0.407	0.190	0.76 (0.77, 1.23)	-2.2	0.87 (0.86, 1.14)	-1.9
72	72	1.102	0.195	0.82 (0.77, 1.23)	-1.7	0.83 (0.82, 1.18)	-2.0
73	73	-0.300	0.188	0.95 (0.77, 1.23)	-0.4	1.01 (0.86, 1.14)	0.2
74	74	1.218	0.198	1.14 (0.77, 1.23)	1.2	0.99 (0.81, 1.19)	-0.1
75	75	0.630	0.187	1.01 (0.77, 1.23)	0.1	0.99 (0.85, 1.15)	-0.1
76	76	-2.283	0.291	0.52 (0.77, 1.23)	-5.1	0.87 (0.58, 1.42)	-0.6
77	77	0.047	0.185	0.76 (0.77, 1.23)	-2.3	0.82 (0.87, 1.13)	-2.8
78	78	-0.091	0.186	0.92 (0.77, 1.23)	-0.7	0.96 (0.87, 1.13)	-0.6
79	79	-1.722	0.244	0.69 (0.77, 1.23)	-3.0	0.91 (0.71, 1.29)	-0.5
80	80	0.457	0.186	1.03 (0.77, 1.23)	0.3	1.01 (0.86, 1.14)	0.2
81	81	-0.195	0.187	0.84 (0.77, 1.23)	-1.4	0.93 (0.87, 1.13)	-1.1
82	82	0.561	0.186	0.96 (0.77, 1.23)	-0.4	0.98 (0.86, 1.14)	-0.2
83	83	0.561	0.186	0.97 (0.77, 1.23)	-0.3	0.96 (0.86, 1.14)	-0.5
84	84	1.218	0.198	1.00 (0.77, 1.23)	0.0	0.87 (0.81, 1.19)	-1.4
85	85	0.953	0.192	1.25 (0.77, 1.23)	2.0	1.12 (0.83, 1.17)	1.4
86	86	0.217	0.185	1.13 (0.77, 1.23)	1.1	1.12 (0.87, 1.13)	1.7
87	87	-0.407	0.190	1.22 (0.77, 1.23)	1.8	1.14 (0.86, 1.14)	1.9
88	88	-0.628	0.194	1.08 (0.77, 1.23)	0.7	1.03 (0.85, 1.15)	0.4
89	89	-0.443	0.190	0.97 (0.77, 1.23)	-0.3	0.96 (0.86, 1.14)	-0.6
90	90	0.217	0.185	0.95 (0.77, 1.23)	-0.4	1.01 (0.87, 1.13)	0.2
91	91	0.012	0.185	1.06 (0.77, 1.23)	0.5	1.10 (0.87, 1.13)	1.4
92	92	0.149	0.185	1.14 (0.77, 1.23)	1.2	1.15 (0.87, 1.13)	2.1
93	93	0.665	0.187	1.21 (0.77, 1.23)	1.7	1.19 (0.85, 1.15)	2.3
94	94	0.630	0.187	0.89 (0.77, 1.23)	-1.0	0.92 (0.85, 1.15)	-1.1
95	95	0.149	0.185	1.00 (0.77, 1.23)	0.0	1.02 (0.87, 1.13)	0.2
96	96	0.701	0.188	0.92 (0.77, 1.23)	-0.6	0.92 (0.85, 1.15)	-1.1
97	97	-0.091	0.186	1.07 (0.77, 1.23)	0.7	1.11 (0.87, 1.13)	1.6
98	98	0.354	0.185	0.90 (0.77, 1.23)	-0.8	0.96 (0.86, 1.14)	-0.5
99	99	0.916	0.191	1.10 (0.77, 1.23)	0.9	1.06 (0.84, 1.16)	0.7
100	100	0.012	0.185	1.09 (0.77, 1.23)	0.8	1.06 (0.87, 1.13)	0.8

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.964

Chi-square test of parameter equality = 2259.93, df = 99, Sig Level = 0.000

^ Quick standard errors have been used

Kết quả phân tích sự phù hợp với mô hình của Đề 5

ConQuest: Generalised Item Response Modelling Software Tue Feb 14 12:05 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR [^]	MNSQ	CI	T	MNSQ	CI	T
1 1	-4.270	0.593	1.42 (0.77, 1.23)	3.3	0.98 (0.00, 2.05)	0.1		
2 2	-1.484	0.219	2.49 (0.77, 1.23)	9.3	1.15 (0.77, 1.23)	1.3		
3 3	-3.353	0.400	0.42 (0.77, 1.23)	-6.5	0.96 (0.36, 1.64)	-0.0		
4 4	-3.524	0.429	0.27 (0.77, 1.23)	-9.2	0.81 (0.30, 1.70)	-0.4		
5 5	-4.270	0.593	0.32 (0.77, 1.23)	-8.3	0.95 (0.00, 2.05)	0.1		
6 6	-0.964	0.199	1.41 (0.77, 1.23)	3.2	1.18 (0.82, 1.18)	1.9		
7 7	-1.390	0.215	1.05 (0.77, 1.23)	0.5	1.07 (0.78, 1.22)	0.6		
8 8	0.275	0.184	1.16 (0.77, 1.23)	1.3	1.14 (0.86, 1.14)	1.8		
9 9	-3.724	0.466	0.24 (0.77, 1.23)	-9.8	0.81 (0.22, 1.78)	-0.4		
10 10	-2.460	0.287	1.88 (0.77, 1.23)	6.2	0.92 (0.61, 1.39)	-0.4		
11 11	-0.771	0.194	1.22 (0.77, 1.23)	1.8	1.13 (0.84, 1.16)	1.5		
12 12	-3.203	0.376	0.49 (0.77, 1.23)	-5.5	0.96 (0.41, 1.59)	-0.0		
13 13	-0.886	0.197	1.03 (0.77, 1.23)	0.3	1.04 (0.83, 1.17)	0.5		
14 14	-3.203	0.376	1.83 (0.77, 1.23)	5.8	0.80 (0.41, 1.59)	-0.6		
15 15	-3.524	0.429	0.27 (0.77, 1.23)	-9.2	0.79 (0.30, 1.70)	-0.5		
16 16	-1.436	0.217	1.39 (0.77, 1.23)	3.0	1.03 (0.78, 1.22)	0.3		
17 17	-2.161	0.261	1.05 (0.77, 1.23)	0.5	0.82 (0.67, 1.33)	-1.1		
18 18	-4.270	0.593	0.28 (0.77, 1.23)	-8.9	0.91 (0.00, 2.05)	0.0		
19 19	-3.203	0.376	0.35 (0.77, 1.23)	-7.6	0.82 (0.41, 1.59)	-0.5		
20 20	-1.004	0.200	0.98 (0.77, 1.23)	-0.1	1.02 (0.82, 1.18)	0.2		
21 21	-1.169	0.206	0.96 (0.77, 1.23)	-0.4	0.91 (0.81, 1.19)	-0.9		
22 22	0.619	0.187	0.99 (0.77, 1.23)	-0.0	1.02 (0.84, 1.16)	0.3		
23 23	0.309	0.184	0.79 (0.77, 1.23)	-1.9	0.85 (0.85, 1.15)	-2.2		
24 24	-1.255	0.209	0.72 (0.77, 1.23)	-2.6	0.92 (0.80, 1.20)	-0.8		
25 25	-0.480	0.188	0.94 (0.77, 1.23)	-0.5	0.99 (0.85, 1.15)	-0.1		
26 26	-1.086	0.203	0.74 (0.77, 1.23)	-2.5	0.86 (0.81, 1.19)	-1.5		
27 27	-2.231	0.267	1.34 (0.77, 1.23)	2.7	0.87 (0.65, 1.35)	-0.7		
28 28	0.760	0.189	1.11 (0.77, 1.23)	0.9	1.09 (0.84, 1.16)	1.1		
29 29	-0.098	0.185	1.16 (0.77, 1.23)	1.4	1.09 (0.86, 1.14)	1.3		
30 30	-0.098	0.185	1.31 (0.77, 1.23)	2.5	1.16 (0.86, 1.14)	2.1		
31 31	-0.660	0.192	0.96 (0.77, 1.23)	-0.4	0.98 (0.84, 1.16)	-0.2		
32 32	-0.235	0.185	1.14 (0.77, 1.23)	1.2	1.07 (0.86, 1.14)	1.0		
33 33	-1.633	0.227	0.72 (0.77, 1.23)	-2.7	0.91 (0.75, 1.25)	-0.7		
34 34	-0.235	0.185	0.84 (0.77, 1.23)	-1.5	0.92 (0.86, 1.14)	-1.1		
35 35	0.105	0.184	1.05 (0.77, 1.23)	0.4	1.06 (0.86, 1.14)	0.9		
36 36	0.654	0.188	0.86 (0.77, 1.23)	-1.3	0.90 (0.84, 1.16)	-1.3		
37 37	0.275	0.184	1.11 (0.77, 1.23)	1.0	1.10 (0.86, 1.14)	1.4		
38 38	1.017	0.194	0.85 (0.77, 1.23)	-1.3	0.84 (0.82, 1.18)	-1.9		
39 39	-0.964	0.199	1.00 (0.77, 1.23)	0.0	0.98 (0.82, 1.18)	-0.2		
40 40	-1.127	0.204	0.66 (0.77, 1.23)	-3.3	0.82 (0.81, 1.19)	-2.0		
41 41	-2.635	0.305	0.37 (0.77, 1.23)	-7.3	0.77 (0.57, 1.43)	-1.0		
42 42	0.037	0.184	0.84 (0.77, 1.23)	-1.4	0.89 (0.86, 1.14)	-1.5		
43 43	-0.201	0.185	0.76 (0.77, 1.23)	-2.3	0.84 (0.86, 1.14)	-2.3		
44 44	0.411	0.185	0.97 (0.77, 1.23)	-0.2	0.94 (0.85, 1.15)	-0.8		
45 45	-2.304	0.273	0.63 (0.77, 1.23)	-3.7	0.87 (0.64, 1.36)	-0.7		
46 46	-0.030	0.184	1.00 (0.77, 1.23)	0.0	1.03 (0.86, 1.14)	0.5		
47 47	-0.515	0.189	0.89 (0.77, 1.23)	-1.0	0.98 (0.85, 1.15)	-0.2		
48 48	-1.212	0.207	0.65 (0.77, 1.23)	-3.5	0.83 (0.80, 1.20)	-1.7		

49 49	0.275	0.184	0.93 (0.77, 1.23)	-0.6	0.88 (0.86, 1.14)	-1.7
50 50	-0.847	0.196	0.68 (0.77, 1.23)	-3.1	0.80 (0.83, 1.17)	-2.4
51 51	0.979	0.193	0.95 (0.77, 1.23)	-0.4	0.92 (0.83, 1.17)	-0.9
52 52	-1.212	0.207	0.71 (0.77, 1.23)	-2.7	0.87 (0.80, 1.20)	-1.4
53 53	-0.132	0.185	0.71 (0.77, 1.23)	-2.8	0.79 (0.86, 1.14)	-3.2
54 54	-0.660	0.192	0.69 (0.77, 1.23)	-3.0	0.83 (0.84, 1.16)	-2.3
55 55	-0.339	0.187	0.90 (0.77, 1.23)	-0.9	0.99 (0.86, 1.14)	-0.1
56 56	-0.964	0.199	0.77 (0.77, 1.23)	-2.1	0.88 (0.82, 1.18)	-1.4
57 57	-1.255	0.209	0.67 (0.77, 1.23)	-3.3	0.84 (0.80, 1.20)	-1.6
58 58	0.343	0.185	0.82 (0.77, 1.23)	-1.6	0.87 (0.85, 1.15)	-1.8
59 59	-0.697	0.192	0.67 (0.77, 1.23)	-3.2	0.77 (0.84, 1.16)	-3.0
60 60	0.480	0.186	1.29 (0.77, 1.23)	2.3	1.24 (0.85, 1.15)	2.9
61 61	0.619	0.187	0.91 (0.77, 1.23)	-0.7	0.93 (0.84, 1.16)	-0.9
62 62	-2.460	0.287	0.45 (0.77, 1.23)	-6.1	0.82 (0.61, 1.39)	-0.9
63 63	-0.235	0.185	1.16 (0.77, 1.23)	1.4	1.11 (0.86, 1.14)	1.5
64 64	0.584	0.187	1.02 (0.77, 1.23)	0.2	0.93 (0.85, 1.15)	-0.8
65 65	-0.304	0.186	0.82 (0.77, 1.23)	-1.7	0.88 (0.86, 1.14)	-1.7
66 66	-2.231	0.267	0.43 (0.77, 1.23)	-6.3	0.76 (0.65, 1.35)	-1.4
67 67	-0.697	0.192	0.85 (0.77, 1.23)	-1.3	0.97 (0.84, 1.16)	-0.4
68 68	-1.908	0.243	0.48 (0.77, 1.23)	-5.6	0.77 (0.71, 1.29)	-1.7
69 69	-1.968	0.247	0.70 (0.77, 1.23)	-2.9	0.82 (0.70, 1.30)	-1.2
70 70	0.905	0.192	1.58 (0.77, 1.23)	4.4	1.35 (0.83, 1.17)	3.7
71 71	0.445	0.185	1.18 (0.77, 1.23)	1.5	1.15 (0.85, 1.15)	1.9
72 72	-0.809	0.195	0.81 (0.77, 1.23)	-1.7	0.92 (0.83, 1.17)	-1.0
73 73	-0.886	0.197	0.72 (0.77, 1.23)	-2.7	0.86 (0.83, 1.17)	-1.7
74 74	0.760	0.189	0.91 (0.77, 1.23)	-0.8	0.84 (0.84, 1.16)	-2.1
75 75	0.139	0.184	1.03 (0.77, 1.23)	0.3	1.02 (0.86, 1.14)	0.4
76 76	0.037	0.184	0.93 (0.77, 1.23)	-0.6	0.96 (0.86, 1.14)	-0.5
77 77	1.249	0.200	1.44 (0.77, 1.23)	3.4	1.16 (0.81, 1.19)	1.6
78 78	1.017	0.194	1.34 (0.77, 1.23)	2.7	0.94 (0.82, 1.18)	-0.6
79 79	0.071	0.184	0.94 (0.77, 1.23)	-0.5	1.00 (0.86, 1.14)	-0.0
80 80	-1.086	0.203	0.84 (0.77, 1.23)	-1.4	0.93 (0.81, 1.19)	-0.7
81 81	-0.809	0.195	1.07 (0.77, 1.23)	0.6	1.12 (0.83, 1.17)	1.4
82 82	-0.098	0.185	1.16 (0.77, 1.23)	1.4	1.22 (0.86, 1.14)	2.8
83 83	2.306	0.247	1.40 (0.77, 1.23)	3.1	1.18 (0.71, 1.29)	1.2
84 84	-0.132	0.185	1.05 (0.77, 1.23)	0.4	1.04 (0.86, 1.14)	0.6
85 85	0.105	0.184	1.28 (0.77, 1.23)	2.3	1.22 (0.86, 1.14)	2.9
86 86	1.249	0.200	1.84 (0.77, 1.23)	5.9	1.49 (0.81, 1.19)	4.4
87 87	0.619	0.187	1.25 (0.77, 1.23)	2.1	1.10 (0.84, 1.16)	1.2
88 88	2.306	0.247	1.92 (0.77, 1.23)	6.4	1.10 (0.71, 1.29)	0.7
89 89	0.514	0.186	1.36 (0.77, 1.23)	2.9	1.21 (0.85, 1.15)	2.6
90 90	-0.339	0.187	1.08 (0.77, 1.23)	0.7	1.05 (0.86, 1.14)	0.7
91 91	3.043	0.304	2.54 (0.77, 1.23)	9.5	1.19 (0.58, 1.42)	0.9
92 92	-0.098	0.185	1.25 (0.77, 1.23)	2.0	1.16 (0.86, 1.14)	2.2
93 93	0.584	0.187	1.08 (0.77, 1.23)	0.7	1.00 (0.85, 1.15)	-0.0
94 94	0.584	0.187	1.00 (0.77, 1.23)	0.0	1.02 (0.85, 1.15)	0.3
95 95	1.819	0.221	1.35 (0.77, 1.23)	2.8	1.07 (0.76, 1.24)	0.6
96 96	0.760	0.189	0.96 (0.77, 1.23)	-0.3	0.94 (0.84, 1.16)	-0.7
97 97	-0.064	0.184	1.10 (0.77, 1.23)	0.9	1.07 (0.86, 1.14)	1.0
98 98	0.275	0.184	1.03 (0.77, 1.23)	0.3	1.04 (0.86, 1.14)	0.6
99 99	-0.270	0.186	1.07 (0.77, 1.23)	0.6	1.11 (0.86, 1.14)	1.5
100 100	1.055	0.195	0.95 (0.77, 1.23)	-0.4	0.97 (0.82, 1.18)	-0.3

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.972

Chi-square test of parameter equality = 2717.89, df = 100, Sig Level = 0.000

^ Quick standard errors have been used

Kết quả phân tích sự phù hợp với mô hình của Đề 6

ConQuest: Generalised Item Response Modelling Software Tue Feb 14 12:16 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1 1	-2.312	0.239	1.11 (0.79, 1.21)	1.0	0.92 (0.73, 1.27)	-0.5		
2 2	-2.492	0.250	1.54 (0.79, 1.21)	4.3	1.06 (0.70, 1.30)	0.4		
3 3	-2.256	0.236	0.82 (0.79, 1.21)	-1.7	0.95 (0.73, 1.27)	-0.3		
4 4	-3.914	0.404	0.98 (0.79, 1.21)	-0.1	0.92 (0.35, 1.65)	-0.1		
5 5	-0.893	0.189	1.27 (0.79, 1.21)	2.3	1.06 (0.84, 1.16)	0.7		
6 6	-1.462	0.202	0.64 (0.79, 1.21)	-3.7	0.87 (0.82, 1.18)	-1.4		
7 7	-4.848	0.597	0.31 (0.79, 1.21)	-8.9	0.93 (0.00, 2.06)	0.0		
8 8	-3.497	0.346	2.33 (0.79, 1.21)	9.0	1.13 (0.48, 1.52)	0.6		
9 9	-3.760	0.381	0.52 (0.79, 1.21)	-5.3	0.88 (0.40, 1.60)	-0.3		
10 10	-2.201	0.233	0.95 (0.79, 1.21)	-0.4	1.02 (0.74, 1.26)	0.2		
11 11	-5.277	0.724	0.42 (0.79, 1.21)	-6.9	1.05 (0.00, 2.33)	0.3		
12 12	-2.915	0.284	2.40 (0.79, 1.21)	9.3	1.07 (0.62, 1.38)	0.4		
13 13	-4.537	0.522	0.52 (0.79, 1.21)	-5.3	0.86 (0.10, 1.90)	-0.2		
14 14	-4.292	0.471	1.81 (0.79, 1.21)	6.0	0.98 (0.21, 1.79)	0.1		
15 15	-2.762	0.271	0.63 (0.79, 1.21)	-3.9	0.97 (0.65, 1.35)	-0.1		
16 16	-2.998	0.291	0.60 (0.79, 1.21)	-4.2	0.89 (0.60, 1.40)	-0.5		
17 17	-1.264	0.196	1.07 (0.79, 1.21)	0.6	1.06 (0.83, 1.17)	0.7		
18 18	-4.537	0.522	1.07 (0.79, 1.21)	0.6	0.88 (0.10, 1.90)	-0.1		
19 19	-3.914	0.404	0.32 (0.79, 1.21)	-8.7	0.85 (0.35, 1.65)	-0.4		
20 20	-2.256	0.236	0.87 (0.79, 1.21)	-1.2	1.04 (0.73, 1.27)	0.3		
21 21	-3.383	0.332	0.99 (0.79, 1.21)	-0.1	0.97 (0.51, 1.49)	-0.0		
22 22	-2.998	0.291	0.42 (0.79, 1.21)	-6.9	0.85 (0.60, 1.40)	-0.7		
23 23	-1.805	0.214	0.80 (0.79, 1.21)	-2.0	1.02 (0.79, 1.21)	0.2		
24 24	-3.277	0.320	0.44 (0.79, 1.21)	-6.6	0.85 (0.54, 1.46)	-0.6		
25 25	-4.292	0.471	0.27 (0.79, 1.21)	-9.7	0.82 (0.21, 1.79)	-0.3		
26 26	-1.672	0.208	1.00 (0.79, 1.21)	0.0	1.06 (0.80, 1.20)	0.6		
27 27	-2.045	0.224	0.53 (0.79, 1.21)	-5.2	0.84 (0.76, 1.24)	-1.4		
28 28	-1.503	0.203	1.02 (0.79, 1.21)	0.2	0.97 (0.81, 1.19)	-0.3		
29 29	-1.672	0.208	0.86 (0.79, 1.21)	-1.3	1.12 (0.80, 1.20)	1.2		
30 30	-4.292	0.471	0.35 (0.79, 1.21)	-8.0	0.85 (0.21, 1.79)	-0.2		
31 31	-1.716	0.210	0.91 (0.79, 1.21)	-0.8	1.10 (0.80, 1.20)	1.0		
32 32	-1.851	0.215	0.63 (0.79, 1.21)	-3.9	0.87 (0.78, 1.22)	-1.2		
33 33	-1.462	0.202	1.45 (0.79, 1.21)	3.6	1.07 (0.82, 1.18)	0.8		
34 34	-2.690	0.265	1.42 (0.79, 1.21)	3.5	1.05 (0.66, 1.34)	0.3		
35 35	-0.858	0.189	1.13 (0.79, 1.21)	1.1	1.11 (0.84, 1.16)	1.2		
36 36	-2.622	0.260	0.51 (0.79, 1.21)	-5.5	0.92 (0.68, 1.32)	-0.4		
37 37	-0.031	0.185	1.06 (0.79, 1.21)	0.6	0.92 (0.83, 1.17)	-1.0		
38 38	-1.188	0.195	0.76 (0.79, 1.21)	-2.3	0.89 (0.83, 1.17)	-1.2		
39 39	-2.915	0.284	0.36 (0.79, 1.21)	-7.9	0.76 (0.62, 1.38)	-1.3		
40 40	-3.277	0.320	1.38 (0.79, 1.21)	3.1	1.02 (0.54, 1.46)	0.2		
41 41	-1.226	0.195	0.62 (0.79, 1.21)	-4.0	0.81 (0.83, 1.17)	-2.3		
42 42	-1.760	0.212	0.83 (0.79, 1.21)	-1.6	1.03 (0.79, 1.21)	0.3		
43 43	-0.371	0.185	1.03 (0.79, 1.21)	0.3	0.85 (0.84, 1.16)	-1.9		
44 44	-2.622	0.260	0.58 (0.79, 1.21)	-4.5	0.90 (0.68, 1.32)	-0.6		
45 45	-0.929	0.190	0.87 (0.79, 1.21)	-1.2	0.87 (0.84, 1.16)	-1.7		
46 46	-0.439	0.185	0.83 (0.79, 1.21)	-1.6	0.84 (0.84, 1.16)	-1.9		
47 47	-3.383	0.332	0.65 (0.79, 1.21)	-3.7	0.94 (0.51, 1.49)	-0.2		
48 48	-3.497	0.346	0.96 (0.79, 1.21)	-0.3	0.89 (0.48, 1.52)	-0.3		

49	49	-1.946	0.220	0.77 (0.79, 1.21)	-2.3	0.95 (0.77, 1.23)	-0.4
50	50	-4.089	0.433	0.38 (0.79, 1.21)	-7.5	0.92 (0.29, 1.71)	-0.1
51	51	-1.805	0.214	0.61 (0.79, 1.21)	-4.1	0.92 (0.79, 1.21)	-0.7
52	52	-0.201	0.184	1.20 (0.79, 1.21)	1.8	1.18 (0.83, 1.17)	2.1
53	53	-1.946	0.220	0.53 (0.79, 1.21)	-5.3	0.82 (0.77, 1.23)	-1.6
54	54	-0.371	0.185	1.41 (0.79, 1.21)	3.4	1.34 (0.84, 1.16)	3.7
55	55	-2.555	0.255	0.56 (0.79, 1.21)	-4.8	0.85 (0.69, 1.31)	-0.9
56	56	-1.851	0.215	1.00 (0.79, 1.21)	0.0	0.98 (0.78, 1.22)	-0.1
57	57	-0.543	0.186	1.19 (0.79, 1.21)	1.7	1.14 (0.84, 1.16)	1.7
58	58	-0.822	0.188	1.17 (0.79, 1.21)	1.5	1.17 (0.84, 1.16)	2.0
59	59	0.072	0.185	1.37 (0.79, 1.21)	3.1	1.18 (0.83, 1.17)	2.0
60	60	-0.474	0.185	0.99 (0.79, 1.21)	-0.0	0.95 (0.84, 1.16)	-0.6
61	61	-4.537	0.522	2.01 (0.79, 1.21)	7.2	0.89 (0.10, 1.90)	-0.1
62	62	-1.226	0.195	1.06 (0.79, 1.21)	0.6	1.04 (0.83, 1.17)	0.5
63	63	0.702	0.190	0.85 (0.79, 1.21)	-1.5	0.88 (0.82, 1.18)	-1.4
64	64	-0.235	0.184	1.10 (0.79, 1.21)	0.9	1.02 (0.83, 1.17)	0.2
65	65	-0.337	0.185	0.66 (0.79, 1.21)	-3.6	0.76 (0.84, 1.16)	-3.1
66	66	-3.497	0.346	0.48 (0.79, 1.21)	-5.9	0.91 (0.48, 1.52)	-0.3
67	67	0.004	0.185	1.18 (0.79, 1.21)	1.6	1.14 (0.83, 1.17)	1.5
68	68	-4.089	0.433	0.41 (0.79, 1.21)	-7.1	0.86 (0.29, 1.71)	-0.3
69	69	-3.497	0.346	0.66 (0.79, 1.21)	-3.6	1.02 (0.48, 1.52)	0.2
70	70	-1.760	0.212	0.94 (0.79, 1.21)	-0.5	0.98 (0.79, 1.21)	-0.1
71	71	-1.629	0.207	0.77 (0.79, 1.21)	-2.3	0.95 (0.80, 1.20)	-0.4
72	72	0.140	0.185	1.27 (0.79, 1.21)	2.3	1.17 (0.83, 1.17)	1.9
73	73	-2.555	0.255	0.81 (0.79, 1.21)	-1.8	1.02 (0.69, 1.31)	0.2
74	74	-0.508	0.185	0.87 (0.79, 1.21)	-1.2	0.94 (0.84, 1.16)	-0.7
75	75	-1.150	0.194	0.78 (0.79, 1.21)	-2.2	0.91 (0.83, 1.17)	-1.1
76	76	-1.075	0.192	0.84 (0.79, 1.21)	-1.5	0.96 (0.83, 1.17)	-0.5
77	77	-1.075	0.192	0.66 (0.79, 1.21)	-3.5	0.84 (0.83, 1.17)	-1.9
78	78	-0.543	0.186	1.04 (0.79, 1.21)	0.4	1.06 (0.84, 1.16)	0.7
79	79	-1.544	0.204	1.75 (0.79, 1.21)	5.7	1.17 (0.81, 1.19)	1.7
80	80	-1.946	0.220	0.73 (0.79, 1.21)	-2.7	0.96 (0.77, 1.23)	-0.3
81	81	0.630	0.189	1.18 (0.79, 1.21)	1.6	0.78 (0.82, 1.18)	-2.6
82	82	-0.439	0.185	1.28 (0.79, 1.21)	2.4	1.11 (0.84, 1.16)	1.3
83	83	-1.805	0.214	1.27 (0.79, 1.21)	2.3	1.09 (0.79, 1.21)	0.9
84	84	-1.586	0.205	0.93 (0.79, 1.21)	-0.6	0.96 (0.81, 1.19)	-0.3
85	85	-0.893	0.189	1.01 (0.79, 1.21)	0.1	1.04 (0.84, 1.16)	0.5
86	86	-1.946	0.220	0.71 (0.79, 1.21)	-3.0	0.87 (0.77, 1.23)	-1.2
87	87	-1.805	0.214	0.71 (0.79, 1.21)	-3.0	0.92 (0.79, 1.21)	-0.7
88	88	0.775	0.191	1.50 (0.79, 1.21)	4.0	1.24 (0.82, 1.18)	2.4
89	89	-1.629	0.207	0.99 (0.79, 1.21)	-0.1	1.02 (0.80, 1.20)	0.2
90	90	-0.371	0.185	1.33 (0.79, 1.21)	2.8	1.22 (0.84, 1.16)	2.5
91	91	-3.085	0.300	1.42 (0.79, 1.21)	3.4	1.05 (0.58, 1.42)	0.3
92	92	-0.508	0.185	1.16 (0.79, 1.21)	1.5	1.08 (0.84, 1.16)	0.9
93	93	-1.586	0.205	0.74 (0.79, 1.21)	-2.6	0.93 (0.81, 1.19)	-0.7
94	94	-0.751	0.188	1.24 (0.79, 1.21)	2.1	1.19 (0.84, 1.16)	2.1
95	95	-3.622	0.362	0.70 (0.79, 1.21)	-3.0	0.95 (0.44, 1.56)	-0.1
96	96	-3.383	0.332	0.86 (0.79, 1.21)	-1.4	0.99 (0.51, 1.49)	0.0
97	97	-1.995	0.222	1.64 (0.79, 1.21)	5.0	1.16 (0.77, 1.23)	1.3
98	98	-1.462	0.202	0.98 (0.79, 1.21)	-0.1	1.06 (0.82, 1.18)	0.7
99	99	-0.337	0.185	0.83 (0.79, 1.21)	-1.6	0.92 (0.84, 1.16)	-0.9
100	100	0.004	0.185	1.74 (0.79, 1.21)	5.6	1.47 (0.83, 1.17)	4.8

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.960

Chi-square test of parameter equality = 5947.34, df = 100, Sig Level = 0.000

^ Quick standard errors have been used

Kết quả phân tích sự phù hợp với mô hình của Đề 7

ConQuest: Generalised Item Response Modelling Software Mon Mar 13 15:21 2023

TABLES OF RESPONSE MODEL PARAMETER ESTIMATES

=====

TERM 1: item

VARIABLES		UNWEIGHTED FIT				WEIGHTED FIT		
item	ESTIMATE	ERROR^	MNSQ	CI	T	MNSQ	CI	T
1 1	-2.915	0.350	2.03 (0.78, 1.22)	7.0	0.73 (0.50, 1.50)	-1.1		
2 2	-3.184	0.385	0.59 (0.78, 1.22)	-4.2	0.78 (0.42, 1.58)	-0.7		
3 3	-1.229	0.225	0.64 (0.78, 1.22)	-3.5	0.87 (0.76, 1.24)	-1.1		
4 4	-2.687	0.325	0.97 (0.78, 1.22)	-0.3	0.77 (0.55, 1.45)	-1.0		
5 5	-2.915	0.350	1.83 (0.78, 1.22)	5.9	0.74 (0.50, 1.50)	-1.0		
6 6	-2.687	0.325	1.45 (0.78, 1.22)	3.5	0.73 (0.55, 1.45)	-1.3		
7 7	-3.518	0.436	0.18 (0.78, 1.22)	-11.5	0.76 (0.31, 1.69)	-0.6		
8 8	-3.340	0.407	0.62 (0.78, 1.22)	-3.8	0.87 (0.38, 1.62)	-0.3		
9 9	-2.150	0.278	0.57 (0.78, 1.22)	-4.5	0.76 (0.65, 1.35)	-1.4		
10 10	-2.150	0.278	1.22 (0.78, 1.22)	1.8	0.93 (0.65, 1.35)	-0.3		
11 11	-3.518	0.436	0.47 (0.78, 1.22)	-5.8	0.89 (0.31, 1.69)	-0.2		
12 12	-0.190	0.196	1.43 (0.78, 1.22)	3.4	1.13 (0.83, 1.17)	1.4		
13 13	-3.340	0.407	1.73 (0.78, 1.22)	5.3	0.99 (0.38, 1.62)	0.1		
14 14	-0.717	0.208	1.66 (0.78, 1.22)	4.9	1.24 (0.81, 1.19)	2.3		
15 15	-0.894	0.213	0.72 (0.78, 1.22)	-2.7	0.88 (0.79, 1.21)	-1.2		
16 16	-1.386	0.232	0.69 (0.78, 1.22)	-3.0	0.96 (0.75, 1.25)	-0.2		
17 17	-3.340	0.407	0.20 (0.78, 1.22)	-10.9	0.76 (0.38, 1.62)	-0.7		
18 18	-1.612	0.244	0.68 (0.78, 1.22)	-3.1	0.87 (0.72, 1.28)	-0.9		
19 19	-2.687	0.325	0.57 (0.78, 1.22)	-4.4	0.82 (0.55, 1.45)	-0.8		
20 20	-0.507	0.202	0.79 (0.78, 1.22)	-2.0	0.90 (0.82, 1.18)	-1.1		
21 21	-1.179	0.223	0.94 (0.78, 1.22)	-0.5	0.99 (0.77, 1.23)	-0.0		
22 22	-3.043	0.366	0.48 (0.78, 1.22)	-5.6	0.79 (0.46, 1.54)	-0.7		
23 23	-1.441	0.235	0.52 (0.78, 1.22)	-5.2	0.74 (0.74, 1.26)	-2.2		
24 24	-1.672	0.247	1.27 (0.78, 1.22)	2.2	0.81 (0.72, 1.28)	-1.4		
25 25	-2.398	0.298	0.53 (0.78, 1.22)	-4.9	0.80 (0.61, 1.39)	-1.0		
26 26	1.335	0.198	1.25 (0.78, 1.22)	2.0	1.06 (0.82, 1.18)	0.7		
27 27	-3.043	0.366	2.05 (0.78, 1.22)	7.1	0.86 (0.46, 1.54)	-0.5		
28 28	-1.931	0.263	0.58 (0.78, 1.22)	-4.3	0.87 (0.68, 1.32)	-0.8		
29 29	-0.268	0.198	0.77 (0.78, 1.22)	-2.1	0.88 (0.83, 1.17)	-1.4		
30 30	-0.000	0.194	0.83 (0.78, 1.22)	-1.6	0.92 (0.83, 1.17)	-1.0		
31 31	-0.632	0.205	0.59 (0.78, 1.22)	-4.2	0.74 (0.81, 1.19)	-3.0		
32 32	-2.915	0.350	0.54 (0.78, 1.22)	-4.8	0.81 (0.50, 1.50)	-0.7		
33 33	-2.311	0.290	0.66 (0.78, 1.22)	-3.4	0.74 (0.62, 1.38)	-1.5		
34 34	0.296	0.191	0.93 (0.78, 1.22)	-0.6	0.94 (0.84, 1.16)	-0.7		
35 35	-2.311	0.290	2.27 (0.78, 1.22)	8.3	1.23 (0.62, 1.38)	1.2		
36 36	0.296	0.191	0.83 (0.78, 1.22)	-1.6	0.88 (0.84, 1.16)	-1.4		
37 37	-0.000	0.194	1.56 (0.78, 1.22)	4.2	1.26 (0.83, 1.17)	2.9		
38 38	-2.229	0.284	0.44 (0.78, 1.22)	-6.2	0.81 (0.64, 1.36)	-1.1		
39 39	-1.931	0.263	1.77 (0.78, 1.22)	5.5	1.02 (0.68, 1.32)	0.2		
40 40	-0.000	0.194	1.01 (0.78, 1.22)	0.1	1.08 (0.83, 1.17)	1.0		
41 41	-2.687	0.325	0.92 (0.78, 1.22)	-0.7	0.87 (0.55, 1.45)	-0.5		
42 42	-0.717	0.208	0.80 (0.78, 1.22)	-1.8	0.89 (0.81, 1.19)	-1.1		
43 43	-0.507	0.202	0.67 (0.78, 1.22)	-3.2	0.82 (0.82, 1.18)	-2.0		
44 44	0.296	0.191	1.40 (0.78, 1.22)	3.1	1.21 (0.84, 1.16)	2.4		
45 45	-1.553	0.241	0.58 (0.78, 1.22)	-4.3	0.85 (0.73, 1.27)	-1.1		
46 46	-2.229	0.284	0.59 (0.78, 1.22)	-4.2	0.78 (0.64, 1.36)	-1.3		
47 47	-1.129	0.221	0.79 (0.78, 1.22)	-1.9	0.89 (0.77, 1.23)	-0.9		
48 48	-2.311	0.290	0.92 (0.78, 1.22)	-0.7	0.96 (0.62, 1.38)	-0.2		

49 49	0.186	0.192	0.98 (0.78, 1.22)	-0.1	1.03 (0.84, 1.16)	0.4
50 50	-1.931	0.263	1.25 (0.78, 1.22)	2.0	0.87 (0.68, 1.32)	-0.8
51 51	-0.760	0.209	2.41 (0.78, 1.22)	9.0	1.51 (0.80, 1.20)	4.4
52 52	0.479	0.191	1.04 (0.78, 1.22)	0.4	0.96 (0.84, 1.16)	-0.5
53 53	-1.229	0.225	0.78 (0.78, 1.22)	-2.0	1.01 (0.76, 1.24)	0.1
54 54	-0.307	0.198	0.88 (0.78, 1.22)	-1.0	0.97 (0.83, 1.17)	-0.4
55 55	0.993	0.193	1.28 (0.78, 1.22)	2.3	1.15 (0.83, 1.17)	1.6
56 56	-0.190	0.196	0.93 (0.78, 1.22)	-0.6	1.00 (0.83, 1.17)	-0.0
57 57	-1.229	0.225	0.65 (0.78, 1.22)	-3.4	0.90 (0.76, 1.24)	-0.9
58 58	-0.386	0.200	0.67 (0.78, 1.22)	-3.3	0.82 (0.82, 1.18)	-2.1
59 59	-1.672	0.247	0.56 (0.78, 1.22)	-4.5	0.84 (0.72, 1.28)	-1.1
60 60	-0.307	0.198	0.66 (0.78, 1.22)	-3.4	0.77 (0.83, 1.17)	-2.8
61 61	-0.590	0.204	0.72 (0.78, 1.22)	-2.6	0.81 (0.81, 1.19)	-2.1
62 62	1.143	0.195	1.21 (0.78, 1.22)	1.8	0.90 (0.83, 1.17)	-1.1
63 63	-0.268	0.198	1.03 (0.78, 1.22)	0.3	1.00 (0.83, 1.17)	-0.0
64 64	-1.612	0.244	0.69 (0.78, 1.22)	-3.0	0.88 (0.72, 1.28)	-0.8
65 65	-0.849	0.212	0.65 (0.78, 1.22)	-3.5	0.81 (0.80, 1.20)	-1.9
66 66	0.075	0.193	0.70 (0.78, 1.22)	-2.9	0.81 (0.84, 1.16)	-2.4
67 67	-0.674	0.206	0.78 (0.78, 1.22)	-2.1	0.87 (0.81, 1.19)	-1.4
68 68	-0.717	0.208	0.65 (0.78, 1.22)	-3.5	0.84 (0.81, 1.19)	-1.6
69 69	0.223	0.192	0.87 (0.78, 1.22)	-1.1	0.89 (0.84, 1.16)	-1.4
70 70	-0.075	0.195	0.80 (0.78, 1.22)	-1.9	0.86 (0.83, 1.17)	-1.7
71 71	0.406	0.191	0.93 (0.78, 1.22)	-0.6	0.92 (0.84, 1.16)	-0.9
72 72	1.143	0.195	1.84 (0.78, 1.22)	6.0	1.34 (0.83, 1.17)	3.4
73 73	0.734	0.191	1.46 (0.78, 1.22)	3.6	1.08 (0.83, 1.17)	1.0
74 74	0.333	0.191	0.74 (0.78, 1.22)	-2.5	0.78 (0.84, 1.16)	-2.8
75 75	-0.152	0.196	0.66 (0.78, 1.22)	-3.3	0.78 (0.83, 1.17)	-2.8
76 76	-0.466	0.201	0.74 (0.78, 1.22)	-2.4	0.90 (0.82, 1.18)	-1.1
77 77	0.333	0.191	1.38 (0.78, 1.22)	3.0	1.15 (0.84, 1.16)	1.8
78 78	-0.075	0.195	0.95 (0.78, 1.22)	-0.4	1.01 (0.83, 1.17)	0.1
79 79	-1.798	0.254	0.57 (0.78, 1.22)	-4.4	0.73 (0.70, 1.30)	-1.9
80 80	-0.760	0.209	0.83 (0.78, 1.22)	-1.5	0.94 (0.80, 1.20)	-0.6
81 81	-0.986	0.216	0.98 (0.78, 1.22)	-0.1	1.04 (0.79, 1.21)	0.4
82 82	0.808	0.192	0.84 (0.78, 1.22)	-1.4	0.91 (0.83, 1.17)	-1.1
83 83	-0.307	0.198	0.87 (0.78, 1.22)	-1.2	0.97 (0.83, 1.17)	-0.3
84 84	0.037	0.193	1.22 (0.78, 1.22)	1.8	1.18 (0.83, 1.17)	2.0
85 85	0.075	0.193	1.26 (0.78, 1.22)	2.1	1.17 (0.84, 1.16)	1.9
86 86	0.296	0.191	0.86 (0.78, 1.22)	-1.2	0.96 (0.84, 1.16)	-0.4
87 87	0.552	0.191	1.23 (0.78, 1.22)	1.9	1.16 (0.84, 1.16)	1.9
88 88	-0.000	0.194	1.05 (0.78, 1.22)	0.5	1.06 (0.83, 1.17)	0.8
89 89	-0.152	0.196	1.51 (0.78, 1.22)	3.9	1.33 (0.83, 1.17)	3.5
90 90	-0.346	0.199	1.05 (0.78, 1.22)	0.5	1.14 (0.82, 1.18)	1.5
91 91	-0.632	0.205	0.96 (0.78, 1.22)	-0.3	1.08 (0.81, 1.19)	0.9
92 92	0.406	0.191	1.28 (0.78, 1.22)	2.3	1.14 (0.84, 1.16)	1.6
93 93	0.442	0.191	1.19 (0.78, 1.22)	1.6	1.12 (0.84, 1.16)	1.4
94 94	0.881	0.192	2.09 (0.78, 1.22)	7.3	1.66 (0.83, 1.17)	6.4
95 95	1.105	0.194	1.65 (0.78, 1.22)	4.8	1.14 (0.83, 1.17)	1.5
96 96	-0.268	0.198	1.34 (0.78, 1.22)	2.7	1.26 (0.83, 1.17)	2.8
97 97	-0.229	0.197	1.00 (0.78, 1.22)	0.0	1.11 (0.83, 1.17)	1.3
98 98	0.406	0.191	1.45 (0.78, 1.22)	3.5	1.25 (0.84, 1.16)	2.8
99 99	0.808	0.192	1.29 (0.78, 1.22)	2.4	1.11 (0.83, 1.17)	1.2
100 100	0.075	0.193	1.51 (0.78, 1.22)	3.9	1.35 (0.84, 1.16)	3.7

An asterisk next to a parameter estimate indicates that it is constrained

Separation Reliability = 0.963

Chi-square test of parameter equality = 2660.00, df = 100, Sig Level = 0.000

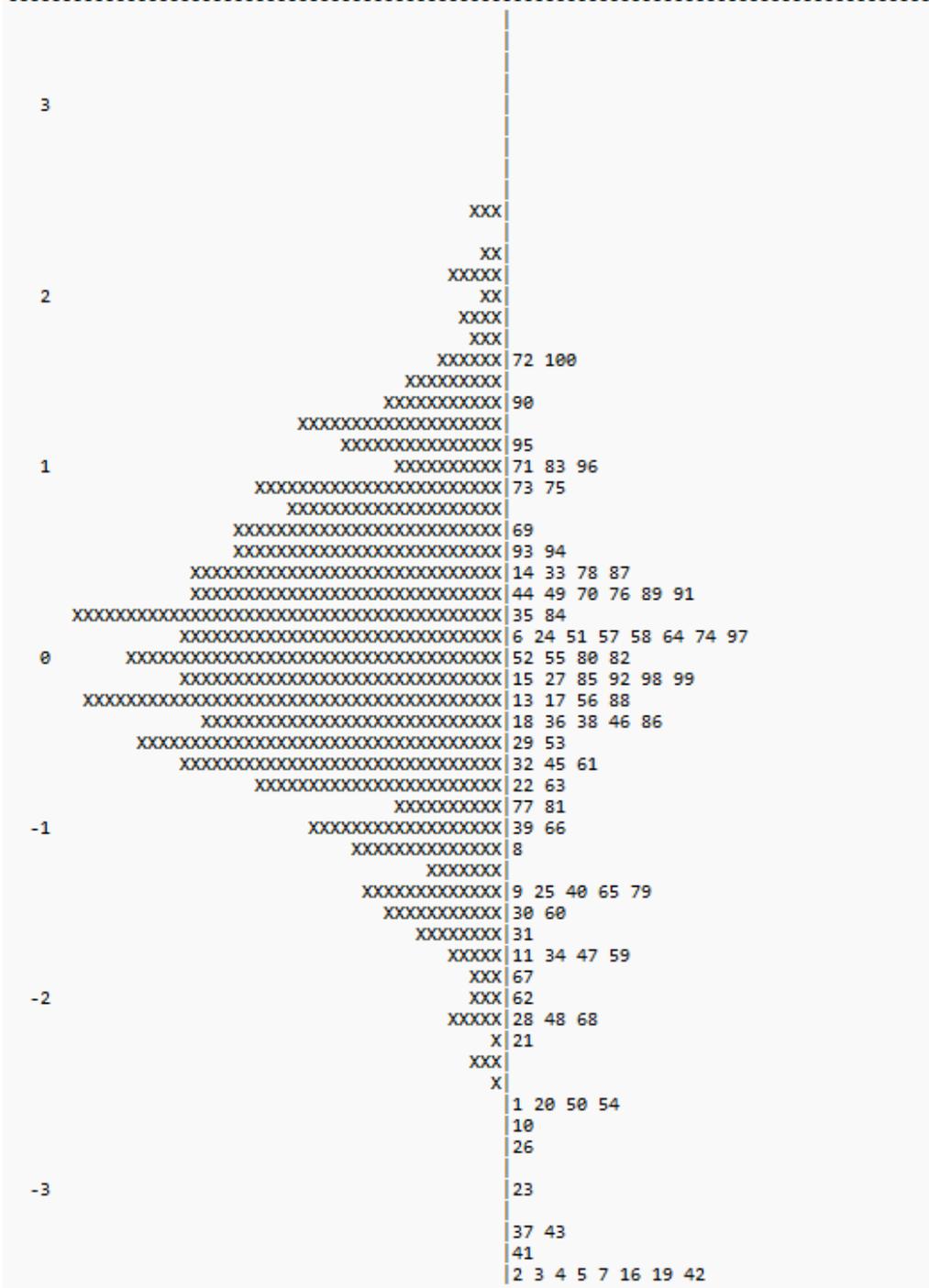
^ Quick standard errors have been used

Bản đồ phân bố năng lực và độ khó của Đề 2

ConQuest: Generalised Item Response Modelling Software Fri Feb 03 21:03 2023
 MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES

Terms in the Model (excl Step terms)

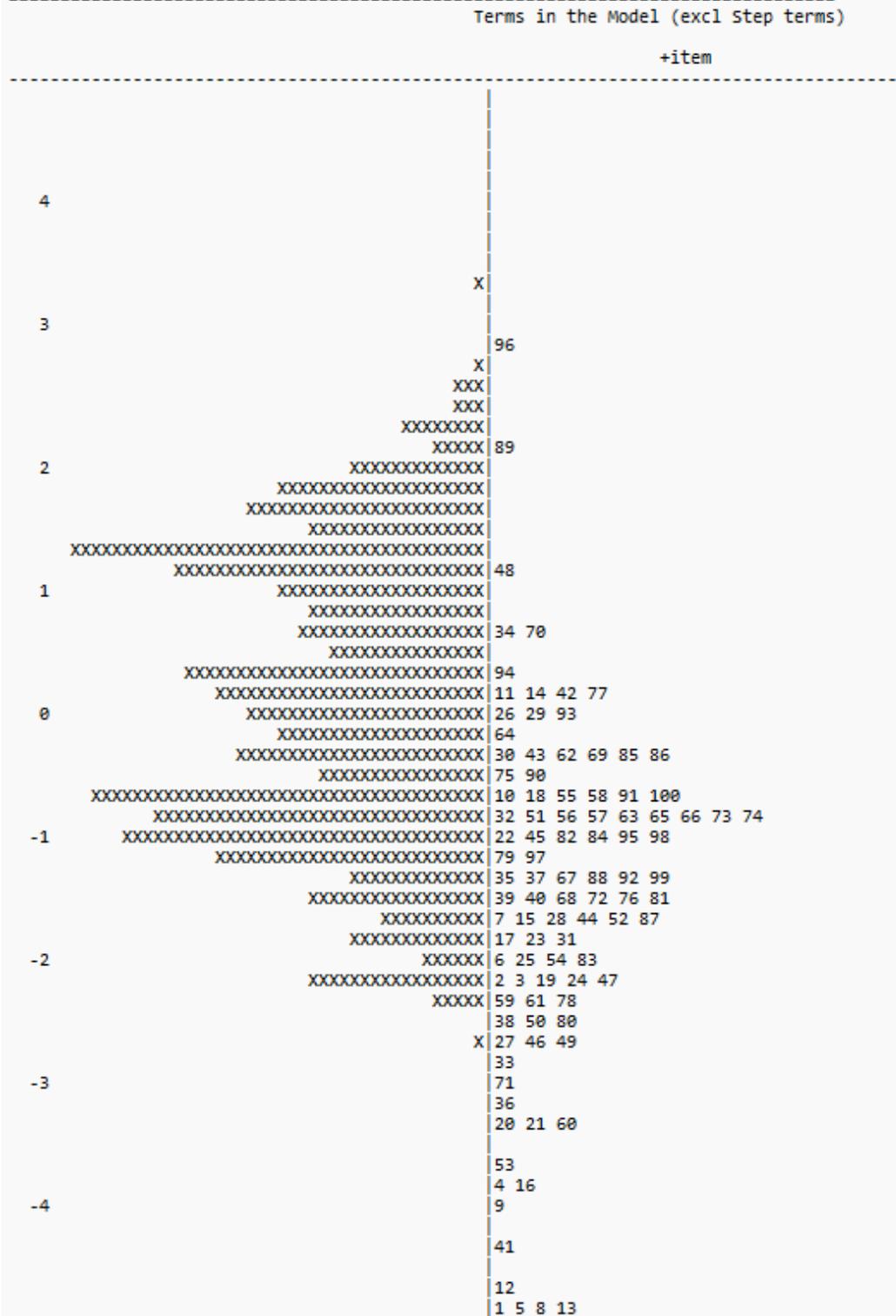
+item



Each 'X' represents 0.2 cases

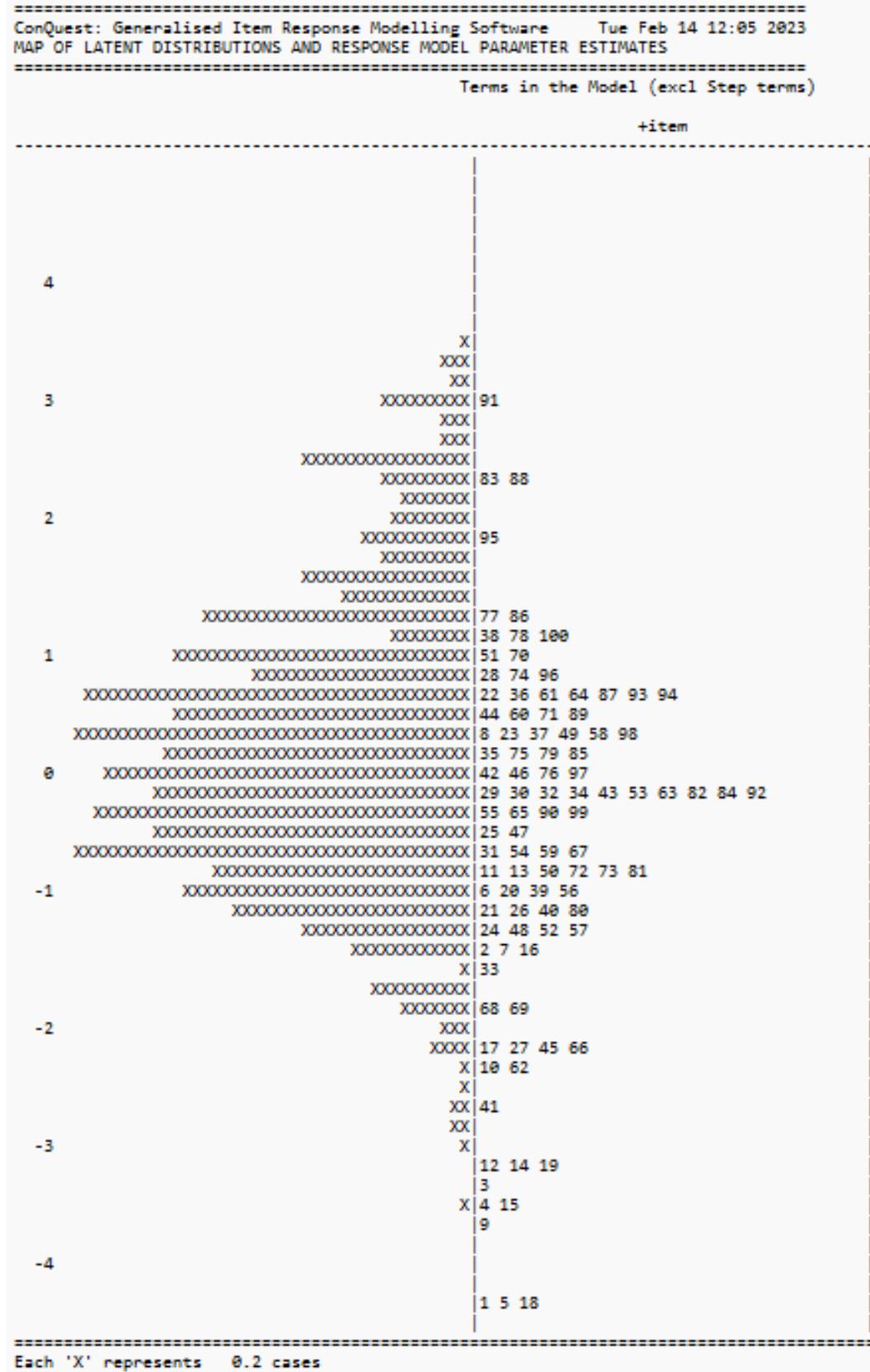
Bản đồ phân bố năng lực và độ khó của Đề 3

ConQuest: Generalised Item Response Modelling Software Sun Feb 05 10:24 2023
 MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES



Each 'X' represents 0.3 cases

Bản đồ phân bố năng lực và độ khó của Đề 5

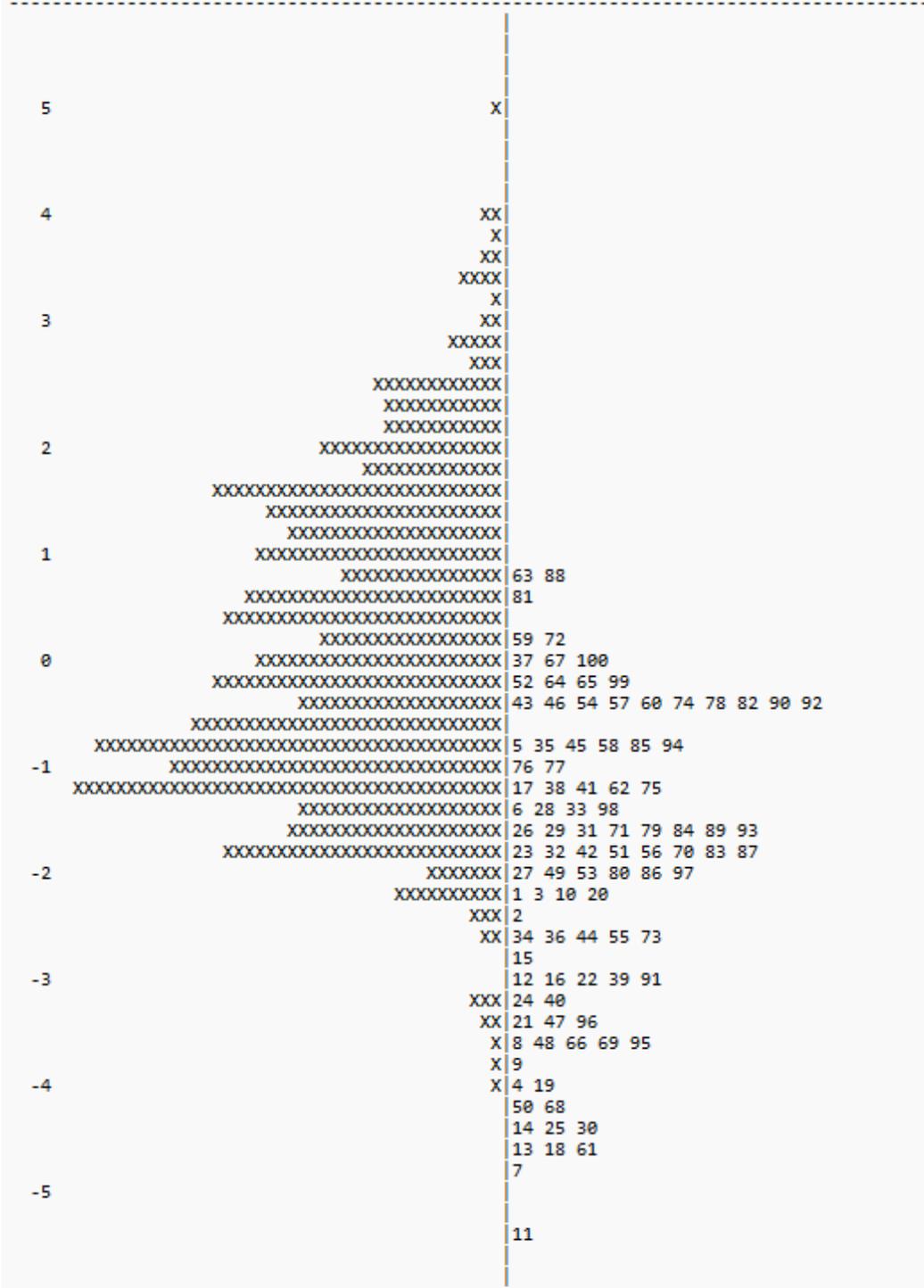


Bản đồ phân bố năng lực và độ khó của Đề 6

ConQuest: Generalised Item Response Modelling Software Tue Feb 14 12:16 2023
 MAP OF LATENT DISTRIBUTIONS AND RESPONSE MODEL PARAMETER ESTIMATES

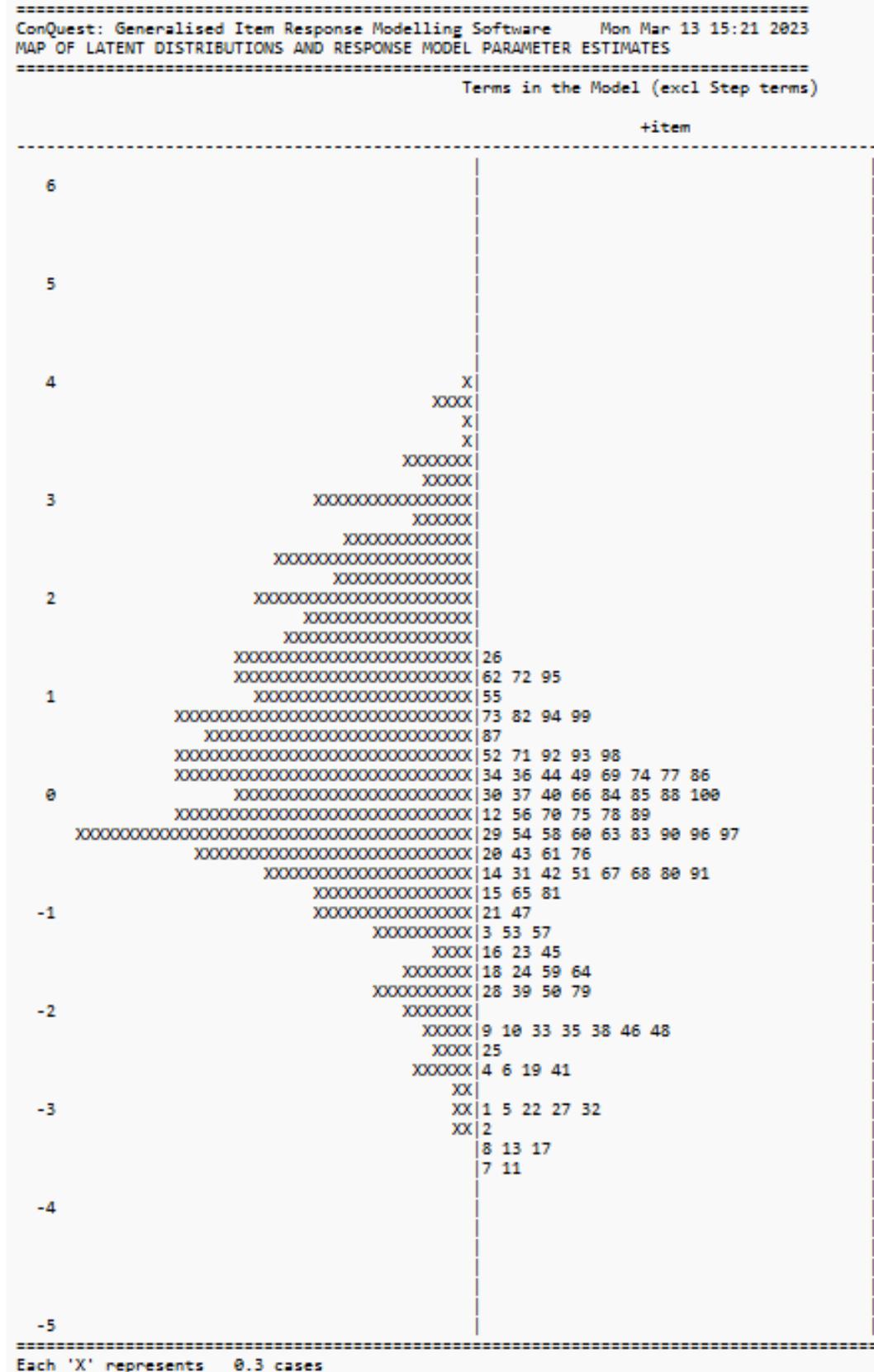
Terms in the Model (excl Step terms)

+item



Each 'X' represents 0.3 cases

Bản đồ phân bố năng lực và độ khó của Đề 7



**Phụ lục 8: Ngân hàng câu hỏi đã chuẩn hóa nhập
trên hệ thống UEd-CAT**

Vi lý do bảo mật, ngân hàng câu hỏi không được trình bày trong luận án, mã QR đường dẫn ngân hàng câu hỏi sau chỉ cho phép truy cập ngân hàng câu hỏi phục vụ quy trình đánh giá luận án.



Phụ lục 9: Báo cáo thử nghiệm trên hệ thống UEd-CAT

STT	Thời gian (s)	Điểm	Số câu hỏi	Số câu trả lời đúng	% trả lời đúng
#1	201	99.92	16	16	100
#2	188	97.28	20	15	75
#3	217	67.98	19	10	53
#4	211	84.18	20	14	70
#5	177	67.51	11	6	55
#6	240	87.26	20	15	75
#7	450	99.85	16	15	94
#8	228	99.63	10	8	80
#9	520	99.97	17	15	88
#10	402	99.91	20	19	95
#11	76	25.62	11	4	36
#12	222	42.62	20	9	45
#13	203	59.6	20	10	50
#14	153	67.51	11	6	55
#15	199	99.98	12	12	100
#16	576	99.99	17	16	94
#17	390	99.93	20	19	95
#18	400	96.96	20	14	70
#19	143	78.11	20	10	50
#20	117	73.99	16	9	56
#21	158	62.82	20	11	55
#22	273	41.33	19	9	47
#23	723	98.64	20	17	85
#24	358	99.64	17	16	94
#25	289	99.92	16	16	100
#26	236	47.49	20	9	45
#27	303	67.26	20	11	55
#28	243	99.92	19	17	89
#29	319	99.59	20	17	85
#30	258	72.77	20	9	45
#31	286	85.86	20	12	60
#32	271	90.3	20	13	65
#33	239	99.85	16	15	94
#34	122	63.65	20	11	55
#35	171	82.95	20	13	65
#36	214	77.31	20	12	60
#37	217	61.49	20	11	55

#38	195	99.92	16	16	100
#39	255	99.98	12	12	100
#40	258	99.97	14	13	93
#41	249	99.93	16	16	100
#42	223	99.98	16	15	94
#43	400	97.75	19	16	84
#44	62	99.98	12	12	100
#45	163	99.98	12	12	100
#46	245	99.98	12	12	100
#47	305	99.92	16	16	100
#48	184	99.98	17	16	94
#49	170	98.81	20	16	80
#50	257	91.5	20	10	50
#51	259	87.12	19	11	58
#52	428	94.92	20	14	70
#53	211	88.04	18	10	56
#54	224	75.68	20	13	65
#55	207	84.3	20	13	65
#56	188	99.19	17	15	88
#57	111	99.98	17	16	94
#58	318	99.98	17	16	94
#59	350	99.02	20	16	80
#60	36	27.36	8	2	25
#61	273	99.97	15	14	93
#62	158	99.92	16	16	100
#63	339	96.83	20	16	80
#64	263	99.22	20	17	85
#65	309	99.98	13	13	100
#66	355	99.71	20	18	90
#67	50	12.8	6	1	17
#68	56	99.97	12	12	100
#69	118	99.91	14	12	86
#70	197	99.97	12	12	100
#71	204	99.85	16	15	94
#72	177	68.76	20	11	55
#73	208	90.91	20	14	70
#74	209	61.91	16	8	50
#75	183	95.7	20	12	60
#76	269	97.38	20	15	75
#77	293	99.46	20	17	85

#78	253	32.67	16	7	44
#79	189	99.69	17	13	76
#80	213	92.65	20	11	55
#81	410	96.91	20	13	65
#82	303	99.84	20	18	90
#83	177	83.1	20	13	65
#84	160	58.71	20	10	50
#85	368	79.87	20	12	60
#86	431	59.44	20	10	50
#87	202	89.54	20	13	65
#88	415	99.76	20	16	80
#89	222	99.95	12	11	92
#90	244	99.93	16	16	100
#91	182	43.76	20	10	50
#92	489	99.98	19	18	95
#93	246	97.08	20	15	75
#94	153	78.96	17	10	59
#95	149	55.13	20	11	55
#96	138	30.85	20	10	50
#97	281	99.92	16	16	100
#98	145	49.52	20	9	45
#99	134	68.76	19	11	58
#100	97	57.57	13	7	54
#101	195	59.95	20	10	50
#102	121	17.64	13	4	31
#103	49	99.84	10	9	90
#104	143	99.97	12	12	100
#105	55	99.97	12	12	100
#106	122	99.97	12	12	100
#107	198	99.97	12	12	100
#108	192	99.85	16	15	94
#109	31	14.01	13	3	23
#110	224	96.84	20	13	65
#111	303	96.76	20	15	75
#112	407	94.06	20	12	60
#113	236	92.38	20	12	60
#114	243	87.41	20	11	55
#115	289	90.41	20	12	60
#116	242	99.96	19	18	95
#117	273	96.18	20	16	80

#118	452	99.92	19	16	84
#119	352	99.85	16	15	94
#120	32	7.47	10	2	20
#121	21	15.48	8	2	25
#122	119	99.98	13	13	100
#123	249	99.55	20	18	90
#124	251	99.32	16	12	75
#125	390	98.67	17	12	71
#126	301	99.49	20	18	90
#127	254	94.6	20	15	75
#128	207	84.73	20	12	60
#129	110	92.41	20	12	60
#130	163	84.38	20	11	55
#131	153	86.54	20	13	65
#132	226	99.39	18	13	72
#133	164	99.92	16	16	100
#134	12	7.61	6	0	0
#135	23	3.68	11	2	18
#136	134	40.28	20	8	40
#137	221	71.33	20	12	60
#138	237	70.3	20	12	60
#139	180	72.64	20	11	55
#140	198	96.61	20	12	60
#141	234	93.19	20	14	70
#142	337	99.14	20	16	80
#143	163	82.04	20	13	65
#144	144	79.3	20	14	70
#145	207	69.6	20	11	55
#146	53	46.14	15	6	40
#147	116	87.7	20	13	65
#148	167	73.57	20	9	45
#149	263	75.8	19	9	47
#150	199	88.18	20	12	60
#151	346	99.48	20	17	85
#152	173	73.5	19	8	42
#153	343	98.13	20	16	80
#154	227	95.03	20	15	75
#155	480	99.98	20	19	95
#156	204	98.58	15	11	73
#157	343	92.13	20	14	70

#158	132	79.47	20	12	60
#159	256	84.92	20	11	55
#160	210	79.33	20	13	65
#161	59	99.98	12	12	100
#162	38	99.98	12	12	100
#163	103	99.98	12	12	100
#164	184	99.92	16	16	100
#165	274	73.63	20	9	45
#166	244	85.5	20	14	70
#167	267	68.93	20	13	65
#168	273	99.85	20	18	90
#169	226	96.24	20	13	65
#170	327	99.92	20	18	90
#171	112	97.35	20	15	75
#172	243	97.24	20	15	75
#173	262	79.16	20	13	65
#174	119	96.52	20	13	65
#175	67	99.95	19	17	89
#176	148	99.19	20	16	80
#177	110	99.95	19	17	89
#178	131	99.95	19	17	89
#179	305	96.77	20	13	65
#180	187	91.15	20	13	65
#181	206	98.19	20	15	75
#182	176	95.72	20	10	50
#183	186	94.58	20	11	55
#184	144	94.7	20	15	75
#185	428	87.9	20	13	65
#186	267	99.28	19	15	79
#187	237	99.55	20	16	80
#188	221	99.57	20	17	85
#189	221	97.26	20	15	75
#190	261	99.92	20	19	95
#191	129	87.45	20	12	60
#192	204	80.32	20	12	60
#193	145	88.18	20	13	65
#194	70	99.98	12	12	100
#195	225	99.97	19	17	89
#196	195	99.92	16	16	100
#197	151	99.92	16	16	100

#198	270	94.32	20	13	65
#199	286	93.52	20	13	65
#200	88	98.63	19	13	68
#201	197	98.83	20	14	70
#202	135	98.74	20	17	85
#203	34	99.98	12	12	100
#204	28	99.98	12	12	100
#205	36	99.98	12	12	100
#206	141	99.94	20	19	95
#207	137	99.98	12	12	100
#208	386	99.97	15	14	93
#209	429	99.92	16	16	100

Phụ lục 10: Nội dung phỏng vấn

Phỏng vấn 1

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Như vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trải nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em xin giới thiệu là em tên là NHẬT, em đang là sinh viên năm nhất chuyên ngành Kinh tế và quản lý.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Cô được biết điểm TOEIC đầu vào của em là 465 đúng không nhỉ?

Vâng đúng ạ.

Vậy là em không tham gia học các khóa học Tiếng Anh cơ sở đúng không?

Vâng ạ

Bây giờ em có tự học tiếng Anh không?

Có nhưng khá ít cô ạ

Em thấy học tiếng Anh thế nào? Sợ gì? Thích gì? Tự tin gì ngại gì?

Em học từ vựng thấy hứng thú hơn so với nghe nói, nghe nói em thấy yếu hơn, ít luyện hơn. Mấy giai đoạn thử nghiệm thì em đều tham gia vào em thấy nó làm cũng không khó lắm. Nó kiểu vừa làm giúp mình ôn lại một số từ vựng để ôn lại kiến thức thì em thấy vui.

Theo em nếu bây giờ thi lại thì điểm tiếng Anh của em sẽ thế nào? Cao hơn thấp hơn hay vẫn vậy?

Nếu thi lại thì điểm sẽ dao động 450-480, vẫn duy trì phong độ vậy, em đoán thế ạ. Em chưa làm nhiều đề TOEIC lắm, lần thi đầu vào là lần thi TOEIC đầu tiên của em ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Cái này hơi khó, cấp 3 em học khá nhiều nhưng không có con số chính xác, em đoán khoảng 800-1200 từ.

Nếu điểm TOEIC là 465 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 2500 từ. Điểm đề hoàn chỉnh của em là 90/100, với danh sách từ 2801 từ thì cũng có ước lượng khoảng 2500 từ đó.

Em ước lượng không chính xác vì không biết làm thế nào, không có công cụ để ước lượng lượng từ em có ạ.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em hiểu trắc nghiệm thích ứng trên máy tính là mình làm đề trải nghiệm tiếng Anh trên một web để thi và ôn các từ vựng, thích ứng ở đây có hữu ích không.

Thích ứng ở đây được hiểu là thích ứng với năng lực của thí sinh. Dựa vào câu trả lời với câu hỏi trước, hệ thống sẽ lựa chọn câu hỏi phù hợp, dễ hơn.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa? Không. Em chưa tham gia bao giờ, đây là lần đầu tiên em có trải nghiệm với trắc nghiệm thích ứng trên máy tính

Em có dùng web hay app gì có hình thức tương tự với trắc nghiệm thích ứng trên máy tính không?

Em có vào web để làm đề và ôn tập tiếng Anh, nhưng cũng chưa có các trải nghiệm với trắc nghiệm thích ứng trên máy tính ạ.

Em có mong đợi gì hay kỳ vọng gì không khi hiểu hơn về trắc nghiệm thích ứng trên máy tính?

Em mong nó giúp em ôn lại từ vựng và mở rộng thêm kiến thức. Em muốn nó không chỉ giúp em kiểm tra lại lượng từ vựng sẵn có mà giúp em học thêm nhiều từ mới.

Em có kỳ vọng thêm về việc mở rộng nội dung với ngữ pháp hay kỹ năng tiếng Anh khác không?

Có, em muốn mở rộng về kỹ năng Nghe, vì em yếu Nghe nên muốn học thêm

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em sử dụng laptop, đa phần là laptop, có một lần dùng điện thoại.

Em có thấy sự khác biệt khi sử dụng các thiết bị khác nhau không?

Em nghĩ dùng laptop sẽ thoải mái hơn, màn hình to, thao tác thoải mái hơn.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trải nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện.

Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em. Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh mà em làm trước đó?

Có, em cảm thấy ngắn hơn.

Có ngắn hơn nhiều hơn, việc hoàn thành với ít thời gian hơn theo em là tốt xấu thế nào? Theo em thời gian đó là hợp lý chưa?

Em thấy tùy câu hỏi dễ khó thì thời gian suy nghĩ sẽ ngắn hoặc dài hơn, đề hoàn chỉnh 100 câu còn đề trắc nghiệm thích ứng trên máy tính thì khoảng 20 câu, thời gian làm bài ngắn hơn nhiều ạ.

Thứ hai là về thao tác thực hiện với trải nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Em thấy thao tác làm trên web đơn giản, không cần kỹ năng máy tính gì đặc biệt, có thể nhiều đối tượng cũng làm được

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá đúng năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Cái này thì em không chắc, em nghĩ em muốn có một bài khó hơn để xem năng lực của mình thế nào. Tuy nhiên, em đã làm ba bài khác nhau với các câu hỏi khác nhau; kết quả em nhận được gần như nhau nên em tin rằng hệ thống đã đánh giá chính xác năng lực của em.

Theo em các câu hỏi có phù hợp với năng lực của thí sinh không?

Em nghĩ là khá phù hợp, với lượng từ 3000 từ thì em thấy phù hợp nhưng em vẫn mong có bài kiểm tra với từ mới và lạ hơn ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Cái này thì không ạ, với mỗi bài em chỉ gặp ít từ khó, các câu còn lại thì dễ và làm khá nhanh, em thấy khá thoải mái.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Khi trắc nghiệm thích ứng trên máy tính lựa chọn câu hỏi phù hợp với năng lực từng thí sinh, thí sinh khác nhau sẽ làm các bài kiểm tra thích ứng trên máy tính khác nhau. Ngay với mỗi thí sinh thì các lượt làm bài cũng khác nhau nên chắc sẽ hạn chế một phần gian lận, vì số lượng câu lặp vẫn có nhưng không nhiều nên không thể nhắc nhau hay nhớ đề.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Có, em nghĩ như thế là hợp lý. Nhưng nếu có phần giải thích đáp án thì mình sẽ biết mình sai ở đâu và sửa ngay lúc đó.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hợp lý không?

Theo em thì khá hợp lý, việc báo kết quả và gợi ý, không đưa đáp án có thể tạo động lực để học hỏi tìm hiểu ạ.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải là một công cụ để em tự đánh giá hiệu quả không?

Một phần ạ, toàn bộ thì hơi khó.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Một công cụ tốt vì làm sai làm lại, làm nhiều nhớ nhiều sẽ tốt ạ.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Em hài lòng nhất là hệ thống dễ truy cập, dễ sử dụng sẽ phù hợp với nhiều đối tượng tham gia. Còn mong đợi em mong các nội dung như ngữ pháp, về nghe sẽ giúp em cải thiện thêm kỹ năng và kiến thức, có thể mở rộng sang các môn học khác có câu hỏi trắc nghiệm, chắc sẽ rất hay vì giúp mình ôn luyện kiến thức.

Có điều gì khiến em không hài lòng và muốn thay đổi không?

Em không thấy có gì khó chịu hay bức xúc gì, chỉ mong khó hơn thì em sẽ học được nhiều hơn.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 2

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em xin giới thiệu là em tên là NHI, em học chuyên ngành Kỹ thuật sinh học ạ.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có e ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không?

Thưa cô, hiện tại vốn tiếng Anh của em không tốt lắm. Khi làm các khảo sát kiểm tra của cô, em thấy được bổ sung kiến thức cho bản thân. Em khá sợ học tiếng Anh.

Trong tất cả những thứ liên quan đến học tiếng Anh, em có cảm thấy tự tin hay e ngại nhất điều gì không?

Em hơi e ngại về phát âm tiếng Anh, em nói tiếng Anh không tốt lắm còn tự tin em thấy tự tin nhất là đọc hiểu, hay áp dụng các công thức ngữ pháp

Cô được biết điểm TOEIC đầu vào của em là 320 đúng không nhỉ?

Vâng đúng ạ.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ điểm thi chắc là cao hơn ạ, em cũng có ý định thi lại trong thời gian tiếp theo ạ, em dự đoán bây giờ mà thi chắc sẽ khoảng 400 ạ, vẫn chưa đạt điều kiện tiếng Anh đầu ra. Em cũng không chắc lắm.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Em nghĩ là không nhiều lắm. Chắc ở khoảng 1500 đến 2000 từ ạ.

Em làm thế nào để có ước lượng đó?

Em không biết cách thức nào, chỉ dự đoán ạ.

Nếu điểm TOEIC là 320 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 1700-2000 từ, em có nghĩ lượng từ của em ở khoảng đó không?

Chắc khoảng đó ạ.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em hiểu trắc nghiệm thích ứng trên máy tính là bài kiểm tra giúp chúng ta tiếp cận dễ hơn với tiếng Anh, hỗ trợ học tiếng Anh dễ dàng hơn.

Em hiểu thích ứng ở đây là gì?

Em hiểu chắc có lẽ là thích ứng gần hơn với những cái chúng ta đang học hiện tại

Thích ứng ở đây được hiểu là thích ứng với năng lực của thí sinh. Dựa vào câu trả lời với câu hỏi trước, hệ thống sẽ lựa chọn câu hỏi phù hợp, dễ hơn hoặc khó hơn từ ngân hàng câu hỏi.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Em chưa tham gia bao giờ, đây là lần đầu tiên em có trải nghiệm với trắc nghiệm thích ứng trên máy tính

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em sử dụng laptop, em làm 2 lượt và đều dùng laptop.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em nghĩ chắc cũng không khác nhau lắm, ví dụ dùng điện thoại thì màn hình nhỏ hơn, có thể ẩn nhầm khi lựa chọn câu hỏi nên với em laptop sẽ là thiết bị tốt nhất để làm bài kiểm tra thích ứng.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Em nghĩ thời gian làm bài là tương đương nhau, tùy thuộc vào câu hỏi, nếu câu hỏi khó thì thời gian suy nghĩ lâu hơn nên thời gian làm bài kiểm tra thích ứng với bài hoàn chỉnh gần như không có sự khác biệt dù số câu hỏi có khác nhau.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Em thấy thao tác làm bài dễ dàng ạ, không yêu cầu kỹ năng máy tính thông thạo hay đặc biệt gì. Em thì cũng làm nhiều bài test trên máy tính vì hồi cấp ba có học online nhiều nên em thấy việc làm bài kiểm tra thích ứng khá là dễ dàng, ngay cả những người chưa làm bài kiểm tra trên máy tính bao giờ thì cũng làm dễ thôi ạ.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em nghĩ cũng tương đối ạ, vì cũng có những câu em chọn đáp án dù không biết gì. Không thể hoàn toàn chính xác vì có thể lựa chọn câu trả lời do đoán mò ạ.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Em không thấy có kiểu một loạt các câu hỏi quá khó hoặc quá dễ khiến em không muốn làm, các câu hỏi có vẻ gần và hợp với năng lực của em ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Dạ không ạ, thấy khá thoải mái. Hiện tại em vẫn muốn trau dồi thêm vốn tiếng Anh nên việc tham gia các bài kiểm tra thế này sẽ giúp mình hơn nên em thấy khá là thoải mái khi làm bài ah.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em nghĩ là bài kiểm tra thế này có tính ưu việt hơn hoặc linh hoạt hơn, có nhiều lựa chọn câu hỏi và đề vì thế sẽ giúp em tăng lượng từ vựng chứ không phải nhớ đáp án cho những lần làm bài sau để có kết quả cao hơn.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Em nghĩ khá là tốt ạ, tại vì đây không phải là bài kiểm tra trên lớp nhưng giúp mình có thể biết mình đang ở đâu, khi nhận được thông báo kết quả ngay lập tức em sẽ nhận được cảm xúc ngay sau khi làm bài, kết quả tốt hay không tốt. Nếu sau một khoảng thời gian nào đó mình đã làm bài đó khá lâu chẳng hạn, mình thậm chí sẽ quên bài đó như thế nào, mình đã trả lời như thế nào. Trước hết là tâm lý, và có thể liên quan đến việc học và ôn luyện của em ạ.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Theo em là có ah. Với mỗi bài như thế em học được từ mới, sử dụng trong các hoàn cảnh khác nhau, mình sẽ có thêm kiến thức về những từ đó. Gợi ý rất hữu ích vì có thể giúp em trước hết là củng cố từ vựng, hai là sẽ biết được nhiều câu hơn, nhiều ngữ cảnh hơn.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải làm một công cụ để em tự đánh giá hiệu quả không?

Em nghĩ cũng có thể ạ. Theo em thì đó là bài kiểm tra có thông báo kết quả ngay, nên sau khi làm bài chúng ta có thể nhìn thấy ngay kết quả luôn, có thể biết năng lực của mình đến đâu, nếu điểm thấp mình có thể làm lại để trau dồi, đến khi điểm cao hơn mình có thể biết năng lực của mình cũng đi lên ạ.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Em nghĩ là có, và nó là công cụ khá là tốt tại vì nếu mà mô hình được nhân rộng hơn thì các bạn sẽ tiếp xúc được với tiếng Anh nhiều hơn, ví dụ như em là một người không học tiếng Anh tốt ngày từ đầu thì nó vừa giúp đánh giá năng lực vừa bổ sung thêm kiến thức để có thể học tiếng Anh một cách dễ hơn.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Em hài lòng nhất là bài kiểm tra này ngắn gọn và rõ ràng, sau khi làm là nhận kết quả ngay, nó rất là trực quan, mình có thể thấy được năng lực của mình và hiệu quả của nó sau khi sử dụng ạ .

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em thì có nhu cầu từ ngữ chuyên ngành nên có đề xuất về mở rộng nội dung ra từ ngữ tiếng Anh chuyên ngành hay các mẫu câu tiếng Anh quen thuộc hay các mẫu câu đặc biệt không thông dụng mà mình hay mắc lỗi.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 3

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em chào cô, em tên là TÂM, em là sinh viên năm thứ nhất, chuyên ngành logistic và quản lý chuỗi cung ứng.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có e ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không?

Lúc trước khi mới vào đại học, em rất ngại, rất sợ học tiếng Anh ạ, em hay quên từ vựng, ngữ pháp còn ỏn hơn một chút. Vào đại học, sau khi thi đầu vào xong, em phải học lại tiếng Anh cơ sở, từ đó em mới bắt đầu học tiếng Anh ạ thì bây giờ đỡ sợ một tí, thấy tự tin và thoải mái hơn.

Cô được biết điểm TOEIC đầu vào của em là 395 đúng không nhỉ?

Vâng đúng ạ, học trong lớp em thấy bình thường, không căng thẳng lắm.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ là sẽ cao hơn rồi đó ạ, em dự định cuối năm em sẽ thi TOEIC, hy vọng đạt điều kiện tiếng Anh đầu ra và duy trì mức đó ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Chắc ở khoảng mấy trăm ạ.

Nếu điểm TOEIC là 395 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 2000 từ, em có nghĩ lượng từ của em ở khoảng đó không?

Em cũng không rõ vì em thấy học ngữ pháp ỏn còn nghe, em không nghe thấy một cái gì, ngữ nghĩa thì em biết, em nhìn thấy từ đó thì em biết nghĩa của từ đó là gì nhưng em không biết nói hay sử dụng từ đó.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em chưa biết gì về trắc nghiệm thích ứng. Khi đang học TACS thì được cô giới thiệu và cho làm bài xây dựng ngân hàng câu hỏi ạ.

Thích ứng ở đây được hiểu là thích ứng với năng lực của thí sinh. Dựa vào câu trả lời với câu hỏi trước, hệ thống sẽ lựa chọn câu hỏi phù hợp, dễ hơn hoặc khó hơn từ ngân hàng câu hỏi.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Đây là trải nghiệm đầu tiên của em. Em mong là được tập luyện thường xuyên để nhớ từ ạ.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em làm bằng laptop lần đầu, lần sau trong giờ nghỉ trên lớp, em dùng điện thoại làm luôn.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em nghĩ là không, chỉ cần có kết nối mạng.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Hình như thời gian cũng như thế ạ, à không, bài trước 100 câu, bài sau 20 câu thì thời gian làm chắc chắn ngắn hơn. Thời gian làm bài kiểm tra thích ứng như thế là vừa đủ với các bạn học tốt và muốn một chút tư duy.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Để thao tác ạ, làm bài kiểm tra như thế thì đa số mọi người cũng làm nhiều rồi.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em cũng không biết có chắc không vì có một số câu em khoanh bừa, không biết vẫn lựa chọn đáp án bất kỳ.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Những câu đầu thì có thể dễ rồi cũng tăng mức độ khó dần lên, mấy câu cuối em không làm được.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Em không, em rất thoải mái, lượng câu hỏi như thế là vừa với em ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em nghĩ là có vì sau khi làm xong câu ấy thì xuất hiện câu mới, không lặp lại nữa, đề em làm các lượt khác nhau cũng khác nhau nên cũng không hỏi được những người thi trước.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Việc thông báo kết quả ngay sau khi làm bài mang lại trải nghiệm rất tích cực, mình biết là mình sai ở đâu, thiếu sót chỗ nào, từ mới là gì, câu ấy được sử dụng thế nào.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Em nghĩ là nó hữu ích, tại vì nó sẽ giúp giải đáp thắc mắc của mình và giúp hiểu hơn câu ấy, có thể làm lại để xem mình đã hiểu chưa. Em cũng khá mong chờ xem nó có giúp mình nhiều hơn không, giúp mình ôn tập lại và tập trung vào những gì mình chưa có.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải là một công cụ để em tự đánh giá hiệu quả không?

Em nghĩ là sẽ giúp em tự đánh giá được năng lực, hoàn toàn thoải mái vì mình em biết kết quả của em.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Chắc chắn là sẽ có, nó không chỉ giúp ích, mà còn giúp ta hiểu biết rộng hơn, biết thêm nhiều hơn. Em muốn sử dụng nó để bổ sung thêm từ vựng.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Theo quan điểm của em, bài kiểm tra khá ổn, đánh giá được năng lực của thí sinh, nhờ đó mình có thể được bổ sung thêm kiến thức, biết thêm nhiều từ mới, cách sử dụng của nó. Em nghĩ thế ah. Em hài lòng nhất là có thể dùng nó để học từ mới.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em nghĩ ứng dụng này khá linh hoạt và tiện lợi nên em cũng không đề xuất để thay đổi gì ạ. Nếu có đề xuất, em muốn thêm tính năng nghe, từ vựng nhưng có thêm bài tập phát âm hay thêm bài tập kỹ năng nghe. Em hiện tại đang rất mong phát triển kỹ năng nghe ạ.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 4

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em xin giới thiệu là em tên là TÚ, em là sinh viên năm nhất chuyên ngành Toán Tin, em đang học TACS 2 do em được 345 TOEIC, em dự định sau khi học TACS2 em sẽ học TOEIC. Hiện em có học từ vựng qua app Mochi Mochi, học từ vựng bằng flashcards.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có e ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Em học TACS1 và TACS2 của trường. Em giao tiếp kém lắm ah, ngại giao tiếp với người nước ngoài, em nghe kém, chỉ học từ vựng ôn thôi ạ.

Trong tất cả những thứ liên quan đến học tiếng Anh, em có cảm thấy tự tin hay e ngại nhất điều gì không?

Em thấy tự tin nhất với từ vựng, ngại giao tiếp ạ.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ sẽ tốt hơn một chút ah, vì học ở trường không định hướng TOEIC, em thấy không hiệu quả. Lớp em đang học đăng ký là 28 bạn mà hôm nay có hơn 10 bạn đi học thôi ạ, không khí học tập không thoải mái lắm, vì thế cũng không hiệu quả ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Theo app em đang học thì có khoảng 2000 từ mà em quên vội nên em nghĩ là khoảng 500 từ em thấy biết chắc chắn trong app, còn tổng lượng từ thì em cũng không rõ vì cũng không có công cụ nào đo ạ. Nếu đọc ra một con số thì chắc hơn 1000 từ ạ

Nếu điểm TOEIC là 345 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 1500-2700 từ, em có nghĩ lượng từ của em ở khoảng đó không? Chắc khoảng 2000 từ ạ, đợt trước em ôn thường xuyên thì chắc khoảng đó, đợt này em đang tập trung ôn thi nhiều môn nên sợ cũng rơi rụng nhiều.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Hôm trước em có vào link làm thì nó dẫn đến một trang web của Đại học Giáo dục, có tìm hiểu chút ít về trắc nghiệm thích ứng.

Thích ứng ở đây được hiểu là thích ứng với năng lực của thí sinh. Dựa vào câu trả lời với câu hỏi trước, hệ thống sẽ lựa chọn câu hỏi phù hợp, dễ hơn hoặc khó hơn từ ngân hàng câu hỏi.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Đây là lần đầu tiên của em ạ. Trước đó em được nghe cô vào lớp giới thiệu về trắc nghiệm thích ứng và xây dựng ngân hàng câu hỏi ạ. Em mong là nội dung từ vựng có nhiều chủ đề, cải thiện web, nếu chuyển thành app thì sẽ tiếp cận được với nhiều đối tượng hơn.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em sử dụng laptop ạ.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em nghĩ smart phone sẽ tiện hơn cho người học, vì laptop chỉ có thể sử dụng ở nhà hoặc cố định ở một số chỗ, điện thoại thì mình có thể học được ở nhiều nơi. Dùng điện thoại mà vào trình duyệt thì không tiện bằng app trên điện thoại.

Cảm ơn em, một gợi ý rất thú vị.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Em thấy như thế là hợp lý rồi ah, thời gian làm bài theo quan điểm của em thì hợp lý ạ. Em thấy làm khoảng 20-30 câu là ổn, nhiều hơn thì làm oải lắm ạ.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Hôm em làm thì thao tác dễ, không có gì khó khăn ạ. Em nghĩ là thao tác thế là dễ, không gây khó khăn hay yêu cầu đòi hỏi kỹ năng gì với người sử dụng ạ.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em nghĩ là có. Nếu chia thành nhiều chủ đề thì sẽ đánh giá tốt hơn ạ.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Theo em là có, khi em làm sai nhiều thì câu hỏi sẽ dễ hơn để mình làm đỡ bị sai hơn, đúng là thích ứng với năng lực thí sinh ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Em không khó chịu gì ạ, cảm thấy ổn, không áp lực thời gian ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Không thể nhớ đề hay nhắc đề. Vì có bạn trả lời nhiều câu đúng hay ít câu đúng thì câu hỏi sau sẽ được đưa ra ngẫu nhiên nên không thể nhớ đề hay nhắc

được ạ. Ví dụ như em làm nhiều lượt thì mỗi lượt câu hỏi lại khác nhau, rất ít gặp lại câu hỏi ạ.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Đúng, em có thấy thông báo kết quả luôn, sẽ có lợi thế hơn việc phải thu bài rồi chấm, kết quả thông báo sau. Việc có kết quả luôn thật sự mang lại trải nghiệm ổn hơn nhiều ạ.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Gợi ý tùy thuộc theo kết quả làm bài của em ạ. Theo em thì gợi ý đó là rất tốt, giúp người học tập trung hơn vào những phần mình chưa tốt ạ.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải là một công cụ để em tự đánh giá hiệu quả không?

Theo em là có ạ. Em thấy nó hữu ích hơn việc chờ đợi việc đánh giá của giáo viên trên lớp ạ.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Em nghĩ nó có thể được ạ. Theo em thì em sẽ sử dụng được trong việc học của em. Ngắn hơn và đơn giản hơn nên nó có thể giúp em ôn tập ạ.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Theo em thì em thấy là bài kiểm tra thích ứng này là phù hợp rồi, nếu mà có nâng cấp thì theo em nên hướng tới từ vựng có chủ đề. Với em, điều em hài lòng nhất là nó dễ thực hiện, dễ thao tác, là một công cụ tốt để ôn tập.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì nữa không?

Em nghĩ là không, nếu có thì nên thêm phần gợi ý cho mỗi câu hỏi, như các gợi
trợ giúp. Nếu mở rộng nội dung, em nghĩ có thể hướng tới ôn tập cấu trúc câu
và luyện tập nói, hay giao tiếp.

**Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em
đã chia sẻ rất nhiệt tình.**

Phỏng vấn 5

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em xin giới thiệu là em tên là NGA, em là sinh viên năm nhất, em học ngành Hóa học, em đang học TACS1 và TACS2 ở trường ạ.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có e ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Em hay học tiếng Anh trên app, đọc truyện xem phim ạ. Học tiếng Anh thì em yếu về nhiều mặt, nghe nói thì rất là yếu, đọc và từ vựng thì khá tự tin, lượng từ vựng của em em nghĩ hỗ trợ tốt cho kỹ năng đọc, em mong phát triển thêm các kỹ năng còn lại.

Cô được biết điểm TOEIC đầu vào của em là 315 đúng không nhỉ?

Vâng đúng ạ.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ sẽ cao hơn, chắc 480-500 vì em thấy thời gian vừa rồi em học tiếng Anh khá hiệu quả ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Em cũng không chắc, chẳng biết ước lượng thế nào, chắc khoảng mấy nghìn từ ạ.

Nếu điểm TOEIC là 315 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 2000 từ, và bây giờ sau một thời gian học có thể tăng thêm nữa đúng không?

Vâng, em cũng nghĩ thế ạ.

Bây giờ chúng ta cùng nói về trải nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trải nghiệm thích ứng trên máy tính nhỉ?

Đây là lần đầu em trải nghiệm về trải nghiệm thích ứng. Từ khi nghe cô giới thiệu về nghiên cứu cô làm, em cũng rất tò mò về trải nghiệm thích ứng ạ. Tuy nhiên đến bài kiểm tra thích ứng từ vựng, em mới có trải nghiệm đầu tiên liên quan đến trải nghiệm thích ứng. Em thấy khá là hay, phù hợp với năng lực của mình, giúp mình biết năng lực của mình đến đâu

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em làm bằng laptop và có sử dụng cả smartphone. Em làm nhiều lượt và sử dụng linh hoạt thiết bị ạ.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em dùng với điện thoại thì khá là ok nhưng khi em dùng laptop thì gặp chút vấn đề.

Em không mở được link làm bài với browser em hay dùng, nó báo có virus. Em phải chuyển sang dùng Cốc Cốc thì mới làm được ah. Còn các thao tác khi làm bài thì không có sự khác biệt giữa laptop và điện thoại ạ.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trải nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Em thấy nó hơi ngắn quá cô ạ, 100 câu thì hơi dài quá, nhưng trên hệ thống có 20 câu thì thấy ngắn quá. Em nghĩ khoảng 40-50 câu thì ổn.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Em không thấy có gì khó, chỉ nhân chọn, rất dễ thao tác, không cần phải có kỹ năng gì đặc biệt, chắc ai cũng làm được.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em nghĩ là có, nó xác định khá là chính xác ạ, em làm khá là ổn và nhận kết quả có vẻ chính xác với năng lực của mình.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Em có thấy. Khi làm đề 100 câu, đoạn đầu em làm nhanh lắm, nhưng đoạn sau phải dừng lại nghĩ. Đến trắc nghiệm thích ứng thì thấy khá là ổn, khi em thấy làm câu khó mà có đáp án sai thì câu hỏi sau lại thấy dễ hơn, khá phù hợp với năng lực của mình.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Em không cô ah, thấy khá thoải mái, chỉ thấy nó ngắn quá nên hơn hụi hẩng.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em thấy điều đó khá là hiệu quả đó ạ. Các lượt làm bài em thấy các câu hỏi gần như là mới, nếu có lặp lại thì cũng là các lượt sau, không phải hai lượt làm bài liên tiếp nên chắc không thể nhớ đề hay nhắc nhau dù cùng làm ạ.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Em thấy em biết luôn kết quả, năng lực của mình đến đâu. So với việc không được nhận kết quả ngay thì việc nhận thông báo kết quả ngay sau khi làm xong là trải nghiệm rất tích cực. Chứ nếu lâu lâu mới nhận được thông báo kết quả, có khi mình quên mất mình đã làm gì và làm thế nào.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Em thấy có hữu ích, em có thể ôn theo ngay những gợi ý đó, và em thấy em làm tốt hơn sau nhiều lượt làm bài.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải làm một công cụ để em tự đánh giá hiệu quả không?

Em nghĩ là có, đánh giá rất là chuẩn ạ. Theo hướng tích cực, việc tự đánh giá sẽ giúp mình có phương án để tự biết mình thế nào, tự điều chỉnh, tự ôn luyện, không cần phụ thuộc vào giáo viên. Nhiều khi trên lớp giáo viên mà đông học sinh ấy thì không thể sát sao với từng cá nhân và đánh giá chuẩn xác được.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Em nghĩ là có, em sẽ rất nhiệt tình sử dụng nó nhiều để tăng từ vựng.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Chắc là nó giúp em biết lượng từ của em đến đâu, em phải ôn tập thêm những gì.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em có mỗi vấn đề là không mở được trong một số trình duyệt. Thêm nữa là, có thêm những lưu ý, định nghĩa cả tiếng Việt với tiếng Anh, đáp án với các câu mình làm sai, từ đó mình có thể học luôn từ đó. Em cũng mong muốn có thêm nhiều trải nghiệm nữa với trắc nghiệm thích ứng trong tương lai, không chỉ với tiếng Anh mà các môn khác cũng sẽ rất hay. Với tiếng Anh thì ngoài từ vựng có thể áp dụng với ngữ pháp, hay từ vựng về chuyên ngành hay từ vựng học thuật ạ.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 6

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em xin giới thiệu là em tên là LAN, em đang là sinh viên chuyên ngành Thực phẩm. Hiện tại học lực tiếng Anh của em cũng khá yếu, từ khi nhập học, em tham gia học TACS1 và hiện giờ đang học TACS2 ạ.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Nói chung là em cũng khá thích học tiếng Anh nhưng cảm thấy học không vào được ạ. Riêng với tiếng Anh em chịu dù cố gắng lắm rồi. Từ vựng của em rất là ít ạ, nó không được nhiều lắm. Thế nên là không thể sử dụng được một cách thành thạo hay giao tiếp. Em thấy nó như không móc nối với nhau.

Cô được biết điểm TOEIC đầu vào của em là 255 đúng không nhỉ?

Vâng đúng ạ.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Không thể nói được ạ, em học mà cứ thấy nó lằng lằng kiểu gì ấy, cảm thấy không tự tin.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Chắc ít lắm ah, có một số từ em học rồi lại quên mất. Em đang thử học với Duolingo. Giờ em chỉ có mục tiêu qua các học phần tiếng Anh em đang học.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em nhớ từ giới thiệu của cô từ hồi trước là thích ứng ở đây được hiểu là thích ứng với năng lực của thí sinh. Dựa vào câu trả lời với câu hỏi trước, hệ thống sẽ lựa chọn câu hỏi dễ hơn hoặc khó hơn để thí sinh làm.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Dạ, bài kiểm tra từ vựng của cô là bài duy nhất, trải nghiệm đầu tiên của em đó ạ. Em mong là nó có phần hồi tích cực trong tương lai.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em toàn bộ dùng máy tính ạ.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Chắc là sẽ không ảnh hưởng gì đâu, dễ thao tác như nhau ạ.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Cái bài vừa rồi là 20 câu ah, theo em thì hơi nhanh một chút ạ. Nếu như với người giỏi thì rất là nhanh. Còn người kém thì cũng mất rất nhiều thời gian, tuy nhiên em mất nhiều thời gian hơn để hoàn thành bài kiểm tra 100 câu, còn bài kiểm tra thích ứng thì thấy hơi nhanh, khoảng 40 câu chắc ổn hơn ạ.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Theo em thao tác khá đơn giản và dễ dàng, không cần phải có kỹ năng máy tính tốt gì cũng có thể làm được ạ.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em nghĩ nhiều khi vẫn có sai lệch ạ, vì khi không biết câu trả lời cứ chọn đại lại có câu trả lời đúng.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Nói chung em thấy khá ổn định, không có kiểu quá nhiều câu dễ hay câu khó liên tục ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Em thấy bình thường vì thời gian khá linh động ạ, bài kiểm tra trên lớp có áp lực thời gian thường mới thấy căng thẳng ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Cứ mỗi một câu lại khác nên gần như không thể nhớ đề nên chắc không có gian lận kiểu nhớ đề nhắc bạn thi sau ạ.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Biết ngay kết quả thì khá ổn nhưng có điều em không biết chính xác đáp án của câu nào mình sai để có thể học luôn.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Nếu như được em muốn biết đáp án những câu mà em làm sai để biết mình sai gì đúng gì và có thể học luôn ạ.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải là một công cụ để em tự đánh giá hiệu quả không?

Cái này thì khó để đảm bảo vì kiểu là mình biết sai câu đó nhưng vẫn không biết đáp án câu đó là gì và có thể học được.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Cái này thì có ạ, mình có thể nhớ được thêm từ vựng và hiểu thêm nên cách học sẽ vui hơn ạ.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Em thấy điểm tốt thì thao tác rất dễ dàng, dễ sử dụng, từ vựng thì thông dụng kiểu dễ nhớ ạ.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Cái chưa phù hợp thì em chưa nghĩ ra, chỉ có mỗi cái em mong đợi là nhìn được đáp án các câu em làm sai để học được luôn ạ. Em rất mong có thêm cách nội dung khác áp dụng trắc nghiệm thích ứng như giao tiếp hàng ngày, các vấn đề về chỉ đường.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 7

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trải nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em xin giới thiệu là em tên là THU, em là sinh viên năm nhất, chuyên ngành Thực phẩm ạ.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Từ khi nhập học, em có học TACS2 và đã hoàn thành vào kỳ trước, hiện em đang học IELTS ở trung tâm bên ngoài ạ. Trong khi học tiếng Anh em thấy ngại nhất là phần Nói và từ vựng. Sau khi em làm các phần thử nghiệm, nếu gặp từ nào quen em sẽ làm được ngay, với các từ mới với em, em sẽ dự đoán thì có cái đúng có cái sai nhưng nhìn chung làm thì ổn. Các từ mới em sẽ ghi ra giấy và học dần học dần, nhờ vậy em biết nhiều từ mới hơn phục vụ cho việc học của em sắp tới ạ.

Trong tất cả những thứ liên quan đến học tiếng Anh, em có cảm thấy tự tin hay ngại nhất điều gì không?

Thực sự thì trong suy nghĩ của em thì một phần em có thích tiếng Anh nhưng nó chỉ là một phần nhỏ, còn một phần lớn là em thấy tiếng Anh rất quan trọng trong học tập và trong công việc sau này của em. Vì vậy, em xác định dù mình có thích hay không cũng phải đặt tiếng Anh lên hàng đầu để nó tốt cho tương lai. Em thích học nhất và tự tin nhất với ngữ pháp, sợ nhất là nói ạ.

Cô được biết điểm TOEIC đầu vào của em là 375 đúng không nhỉ?

Vâng đúng ạ.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ là sẽ có điểm cao hơn. Khi thi đầu vào em còn chưa quen với việc học tiếng Anh cho lắm ạ. Lớp 12 em dành thời gian để ôn thi nhưng không chọn tiếng Anh là môn thi tốt nghiệp của em cho nên đến bây giờ với khả năng của em và thêm thời gian học tập thì kết quả thi của em sẽ phát triển hơn so với kết quả thi đầu vào ạ. Nếu với việc em đang học IELTS hiện tại, em nghĩ mình đang ở mức 4.5 vì trình độ nói của em nó vẫn còn chưa ổn lắm, còn nếu TOEIC thì em nghĩ đã đảm bảo điều kiện tiếng Anh đầu ra ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Chắc khoảng 1000 đến 2000 từ ạ, em không biết lượng từ là bao nhiêu và bằng cách nào nên dự đoán kiểu mò mò vì em tự tin với 1000 từ ạ.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em nghĩ thích ứng là kiểu mình thử nghiệm với nhiều câu hỏi. Em được biết là hệ thống sẽ lựa chọn câu hỏi phù hợp, dễ hơn hoặc khó hơn cho mình làm ạ.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Em chưa ạ. Đây là lần đầu trải nghiệm ạ. Em mong muốn biết được đa dạng từ vựng, lượng từ vựng sẽ được tăng lên.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Thường thì em dùng điện thoại, có lần em dùng laptop.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em cảm thấy không có ảnh hưởng gì ạ, đều có thể đưa ra câu trả lời thoải mái ạ.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Em làm trắc nghiệm thích ứng thấy nhanh hơn nhiều so với 100 câu kia, cũng không gây dễ chán như với đề 100 câu. Khi với nhiều câu thì nhanh chán và nản hơn, nhiều khi làm giữa chừng không muốn làm nữa sẽ out luôn. Với bài kiểm tra thích ứng em thấy không quá dài quá ngắn, số lượng câu hỏi phù hợp.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Không có vấn đề gì thao tác, không gây cản trở gì, khá dễ dàng, không yêu cầu thao tác gì đặc biệt.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em nghĩ là chính xác khoảng 89-90% vì vẫn có từ mà mình không biết thì mình vẫn phải đoán để chọn câu trả lời.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Em nghĩ là phù hợp, trong quá trình làm em không thấy nhiều câu quá dễ hay quá khó, em đủ năng lực để làm ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Dạ không ạ, vì nó không có giới hạn về thời gian nên em khá thoải mái khi làm ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em nghĩ là không, một phần vì làm nhiều lần thì có thể nhớ được từ luôn chứ không cần nhớ đáp án. Em có làm nhiều lần, sau mỗi lần em sẽ lưu lại các từ không biết, dừng lại để tra từ và nhớ từ. Một phần khác thì số câu hỏi nhiều, em gặp nhiều câu hỏi khác nhau mỗi lượt làm, chỉ đôi lần gặp câu hỏi cũ nhưng rất ít và không nhiều nên nhớ đáp án để có kết quả cao trong lượt làm bài sau là gần như không thể ạ.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Em thấy thế là rất ổn, ngay sau khi làm mình biết mình ở mức nào, có thể cải thiện được không. Theo em thì ai cũng thích biết kết quả ngay sau khi làm bài ạ.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Em có thấy các gợi ý ôn lại từ ở mức độ nào hay gì đó nhưng gợi ý hơi ít ạ.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải làm một công cụ để em tự đánh giá hiệu quả không?

Em nghĩ là có khả năng hiệu quả, nó là một phần giúp mình tự đánh giá và ôn tập sau khi học.

Như vậy là em đã đưa cho câu câu trả lời cho câu hỏi sau về việc có thể dùng trắc nghiệm thích ứng trên máy tính như một công cụ học tập và ôn luyện thường xuyên. Cảm ơn em nhé!

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Trước tiên là về từ vựng hữu ích với em, thứ hai là trả lời câu hỏi hay chọn đáp án thì khá dễ dàng, không yêu cầu người làm phải biết sử dụng máy móc các kiểu nên nó hợp với tất cả mọi người. Nếu có thể thì em muốn chèn thêm các hình ảnh để dễ ghi nhớ các từ ạ.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em nghĩ là không, nó không mang lại chút khó chịu ạ, thời gian làm thì nhanh, số câu hỏi thì không quá nhiều nên nhìn chung không có vấn đề gì xảy ra ạ.

Nếu là đề xuất mong muốn thì ngoài việc chèn hình ảnh có thể thêm video hay file ghi âm, việc có thêm hình ảnh âm thanh thú vị sẽ không chỉ thu hút sinh viên mà còn cả các em cấp 1 và cấp 2, thu hút nhiều đối tượng hơn. Ngoài ra có thể đan xen các bài tập từ vựng dễ và khó, các dạng khác nhau không chỉ dừng lại ở ngữ nghĩa, từ đó đòi hỏi hiểu biết của người dùng.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 8

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em thưa cô, em là BÁCH, sinh viên năm nhất chuyên ngành Công nghệ thông tin và truyền thông ạ.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Về học tiếng Anh, em cảm thấy chưa được tốt, đặc biệt về từ vựng, em thấy khá là khó nhớ. Từ vựng là cái em ngại nhất vì khả năng từ vựng yếu nên các phần khác cũng cảm thấy yếu hơn.

Em có thi đầu vào tiếng Anh của trường đúng không?

Điểm TOEIC đầu vào của em là 345 ạ. Em học cả TACS1 và TACS2.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ chỉ ngang điểm đầu vào ạ, có thể hơn chút nhưng chưa đến 400.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Em nghĩ là 200 từ.

Nếu điểm TOEIC là 345 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 2000 từ, em có nghĩ lượng từ của em ở khoảng đó không?

Thực ra em cũng không biết phải ước lượng lượng từ em đang có thể nào ạ, vẫn mong được học thêm nhiều từ nữa ạ.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Những gì cô giới thiệu với em thì đó là những gì em nghe lần đầu về trắc nghiệm thích ứng ạ.

Em hiểu thích ứng ở đây là gì?

Là thích ứng với năng lực người tham gia kiểm tra. Em mong dự án của cô sớm có kết quả và có thể phổ biến rộng rãi đến các bạn sinh viên.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Chưa ạ, đây là trải nghiệm đầu tiên của em ạ.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em sử dụng laptop và cả điện thoại ạ.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em nghĩ cả hai đều có thể sử dụng tốt với trắc nghiệm thích ứng này ạ.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Có ngắn hơn cô ạ, em nghĩ ngắn thế là vừa phải để tập trung vào làm trong một khoảng thời gian. Khi làm bài 100 câu cũng làm khoảng một nửa là muốn dừng, hơi thiếu kiên nhẫn ạ.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Có ạ, rất dễ tiếp cận, dễ sử dụng, không cần dùng kỹ năng máy tính thành thạo, có thể phù hợp với nhiều đối tượng ạ.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Câu này hơi khó ạ. Em nghĩ với em thì đánh giá chính xác năng lực từ vựng em đang có ạ.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Cũng có câu em thấy làm tron tru, cũng có câu em thấy với lượng từ vựng của em không trả lời được. Tuy nhiên, không có nhiều câu liên tiếp quá khó hoặc quá dễ với em ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Lượng từ vựng của em hạn chế nên khi gặp liền mấy câu khó không làm được thì em cũng hơi khó chịu một chút ạ, còn khi làm được thì thấy thoải mái ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em nghĩ với ngân hàng câu hỏi lớn có thể giảm gian lận liên quan đến nhớ đề hay nhắc đề ạ. Em làm mấy lượt thì em không gặp tình trạng lặp lại câu hỏi, lượt sau gặp toàn bộ câu hỏi khác lượt trước ạ.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Đúng ạ. Em thấy đó là một trải nghiệm tốt, nhận kết quả luôn mình biết mình đúng câu nào, phần trăm câu hỏi mình vượt qua là bao nhiêu.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Có, em nghĩ là hữu ích ạ, em biết phần từ vựng kém cần cải thiện. Sau khi nhận gợi ý, em cũng mong có thể học thêm theo các gợi ý đó ạ.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải làm một công cụ để em tự đánh giá hiệu quả không?

Đúng ạ, mình đón nhận kết quả tích cực hơn so với bị đánh giá, mình thấy kết quả và tự thay đổi hay nâng cao năng lực của mình.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Em thấy việc này rất hợp lý ạ. Nếu được sử dụng, mỗi ngày em sẽ dành chút thời gian để vào học từ vựng ạ.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Với em, bài kiểm tra thích ứng mà em trải nghiệm có giao diện thân thiện với người dùng, từ các em nhỏ hơn cũng có thể tiếp cận nó một cách dễ dàng, không cần kỹ năng gì khó cả.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em nghĩ là nên cho thêm kiểm tra từ vựng theo chủ đề mà mình có thể lựa chọn. Cho thêm các lựa chọn về số lượng câu hỏi như bài kiểm tra 20 câu, 30 câu, 50 câu hay 100 câu tùy theo thời gian mà người dùng có.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 9

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trắc nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em là CẨM, K67, sinh viên ngành Toán Tin tại Hà Nội.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có e ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Em khá là thích học tiếng Anh, hồi học cấp 2 ở trường, em học kiểu giải đề nhiều của các kỳ thi, đến cấp 3, em được sử dụng Internet nhiều hơn, em được tiếp cận với các nội dung tiếng Anh nhiều hơn như là âm nhạc, phim, các thứ. Dần dần thì hướng học tiếng Anh của em thêm được hướng giao tiếp bằng cách xem các video âm nhạc và phim bằng tiếng Anh. Tiếng Anh cũng cũng được cải thiện theo hướng khác, không phải chỉ mỗi giải đề.

Trong tất cả những thứ liên quan đến học tiếng Anh, em có cảm thấy tự tin hay e ngại nhất điều gì không?

Em không dám nhận tự tin nhất cái gì bởi vì tiếng Anh của em nó kiểu lợm mỗi thứ một ít, đủ dùng nhưng không bài bản, chưa làm được hết những gì em muốn với tiếng Anh. Còn e ngại nhất thì em e ngại nhất phần nói của mình, vì em nói kiểu bị dính accent nhiều ạ.

Cô được biết điểm TOEIC đầu vào của em là 715 đúng không nhỉ?

Vâng đúng ạ, em không phải tham gia các khóa học tiếng Anh của trường ạ. Em vẫn học và có tham gia câu lạc bộ. Em đã sinh hoạt một lần, trong buổi đó thì em có tham gia các hoạt động. Ngoài ra em cũng vẫn học, mà nói học cũng

không hẳn, em vẫn tiếp tục thu lượm như em vẫn làm chứ không có bài bản gì cả.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Chắc vẫn khoảng đó ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Em nghĩ chắc khoảng trên dưới 1000 một chút ạ, cũng không biết ước lượng thế nào ạ.

Nếu điểm TOEIC là 715 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 3000 từ, em có nghĩ lượng từ của em ở khoảng đó không?

Có thể ạ, các bài thử nghiệm em gần như làm được hết ạ.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em thấy hình như đây là mình trả lời các câu hỏi tiếng Anh để kiểm tra từ vựng của mình ạ.

Em hiểu thích ứng ở đây là gì?

Mỗi lần em làm với một câu hỏi thì câu trả lời của em sẽ quyết định câu hỏi tiếp theo ạ. Em có biết chút ít về trắc nghiệm thích ứng thôi, em nghĩ là nó giúp các bạn kiểm tra có thể biết lượng từ vựng của mình thế nào, em mong nó sẽ ngày càng chính xác hơn, có thể xác định được đúng năng lực của người kiểm tra ạ.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Chưa, đây là lần đầu của em ạ.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em dùng điện thoại ạ.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em nghĩ là không ạ. Em thấy dùng thoải mái với điện thoại.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Em thấy là nó ngắn hơn. Lần trước làm 100 câu, em không tham khảo gì, không tra cứu gì mà làm khá lâu dù cũng làm miệt mài. Còn trắc nghiệm thích ứng thì nó giảm thời lượng đi khá nhiều. Thứ 2, em thấy với 100 câu thì sự phân hóa độ dễ khó không bằng câu hỏi trên hệ thống trắc nghiệm thích ứng, ở hệ thống câu hỏi có vẻ nhiều câu hỏi khó hơn trong số các câu hỏi em làm ạ.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Em thấy khá là dễ dùng, không có gì phức tạp, không có yêu cầu kỹ năng máy tính gì, mình chỉ biết dùng máy tính là được, thao tác rất đơn giản, gần như cơ bản nhất ạ.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Em nghĩ là có, vì theo cơ chế như cô nói, thì nếu mình trả lời đúng sẽ nhận câu hỏi khó hơn. Hệ thống sẽ cho mình câu hỏi khó đến khi mình không trả lời được nữa và xác định được trình độ của mình thì dừng lại, chính cơ chế đấy sẽ mang lại sự chính xác hơn là kiểu mình làm dần trải các câu khó dễ lẫn lộn với nhau, vì khi làm nhiều quá thì nó rất dễ bị loạn và không chính xác lắm. Với cơ chế cô nói thì nó sẽ ok hơn ạ.

Em có cảm thấy cơ chế đó trong bài em làm vừa rồi không?

Em thấy có, vì em cũng cảm nhận thấy câu hỏi khó dần, khó dần lên cho đến khi kết thúc.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Với 20 câu thì các câu hỏi đều gần với năng lực của em, không có các câu dễ quá ạ.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Em không căng thẳng gì lắm, em nghĩ sai thì thôi, vì cũng không phải bài kiểm tra trên lớp.

Nếu dùng trắc nghiệm thích ứng để kiểm tra một nội dung khác như một bài kiểm tra trên lớp thì sao?

Em nghĩ nếu áp dụng vào thi trên lớp thì em nghĩ nó khó mà giảm căng thẳng, vì khi các bạn làm bài mà câu hỏi cứ khó dần lên và sẽ đến mức rất là khó với năng lực của các bạn, thì theo lý thuyết bình thường chắc phải căng thẳng hơn. Nhưng nhìn chung thì nó có thể giảm căng thẳng khi mình không phải làm quá nhiều câu hỏi quá khó hoặc quá dễ so với năng lực của mình ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em nghĩ có thể vì không phải nhiều bạn có cùng trình độ giống nhau, việc trùng đề sẽ hạn chế từ đó hạn chế được gian lận kiểu nhớ đề hay nhớ câu hỏi và nhắc nhau. Ngay cùng năng lực thì cũng gặp đề khác nhau, số lượng câu hỏi sẽ lặp lại khá ít ạ.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Em nghĩ là bạn nào đã tham gia trải nghiệm đó thì tâm lý chung là muốn có kết quả luôn và em nghĩ việc có kết quả luôn như thế là rất là tốt ạ.

Theo em hệ thống ngoài việc thông báo kết quả thì có kèm các gợi ý về học tập không? Các gợi ý ấy có hữu ích hay không?

Em nghĩ rất là hữu ích đó ạ, giúp em học bài đúng hướng hơn, nếu như những bạn có như câu đặc biệt thì lại càng hiệu quả hơn, như kiểu bổ khuyết, bổ sung những cái mình thiếu ạ.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải là một công cụ để em tự đánh giá hiệu quả không?

Em cũng đồng ý ạ. Nó hiệu quả thứ nhất là nó đánh giá đúng lượng từ vựng người thực hiện đang có, và nó hiệu quả và khoa học hơn là mình làm rất nhiều câu hỏi mà thực chất nó không mang lại hiệu quả gì. Nếu theo cách này, mức độ khó tăng dần thì như thế là phản ánh đúng năng lực của người tham gia và nó tiết kiệm được thời gian thay vì mình làm rất nhiều các câu hỏi của bài test. Theo em đó là điểm mạnh của bài này ạ.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Em nghĩ cái này bản chất của nó chỉ là kiểm tra để mình biết mình đang ở mức nào, mình nên làm theo định kỳ, không nên làm thường xuyên. Nó hợp là công cụ đánh giá hơn là công cụ ôn tập.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Em hài lòng nhất với cách mà hệ thống này đã đưa ra những câu hỏi rất là hợp lý để từ đó có thể đánh giá được năng lực của người tham gia, và mất một thời gian ngắn hơn với độ chính xác cao hơn, theo em nghĩ đó là điểm mạnh của hệ thống này.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em chưa nghĩ ra.

Em có mong muốn tham gia trải nghiệm thêm với trắc nghiệm thích ứng trong tương lai không?

Em có ạ, em cũng thích tham gia, thứ nhất là để kiểm tra tiếng Anh của mình, em cũng muốn học hỏi thêm nhiều cái, vì có thể sau này em cũng làm những cái như thế này.

Em có đề xuất gì khác không?

Em nghĩ mình có thể áp dụng với các môn trắc nghiệm khác, không giới hạn về nội dung. Nếu là tiếng Anh thì các câu hỏi ngữ pháp thì chắc không cần lắm, em muốn có các nội dung từ vựng khó hơn như từ vựng phù hợp các chủ đề khác nhau, phục vụ các mục đích khác nhau như từ vựng IELTS, TOEIC.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.

Phỏng vấn 10

Cô đã nhận được cam kết tham gia phỏng vấn của em. Chúng ta bắt đầu cuộc phỏng vấn nhé. Với cuộc phỏng vấn này, câu hỏi sẽ liên quan đến các câu hỏi ở trong khảo sát mà em đã thực hiện. Cô rất mong đợi là em sẽ thể hiện được cái quan điểm rõ ràng hơn, chi tiết hơn, ví dụ như là lý giải tại sao em đồng ý hay phản đối. Nhờ vậy, cô sẽ có nhiều dữ liệu có giá trị và chi tiết hơn để hoàn thành báo cáo nghiên cứu cô đang thực hiện về trải nghiệm thích ứng.

Trước hết, em có thể giới thiệu một chút liên quan đến bản thân em được không?

Em là THANH, sinh viên chuyên ngành Cơ khí ạ.

Bây giờ mình sẽ nói một chút liên quan đến việc học tiếng Anh của em nhé. Em cảm thấy em học tiếng Anh thế nào? Có thích không? Có e ngại hay tự tin về điều gì không? Có muốn thay đổi hay mong muốn điều gì không? Em cảm thấy tiếng Anh với em là khá thú vị, nhưng đến thời điểm này vì nhiều lý do em hơi lười học, thêm vào đó môi trường của em không dùng nhiều tiếng Anh lắm. Em thường học tiếng Anh bằng cách em nghe thụ động ở trên Youtube với các web nghe. Em nghe thụ động trước khi đi ngủ, nhiều khi em nghe bằng cách xem TEDed hoặc một số kênh liên quan, em thấy học những cái tiếng Anh mà có làm hoạt hình thì rất là thú vị.

Trong tất cả những thứ liên quan đến học tiếng Anh, em có cảm thấy tự tin hay e ngại nhất điều gì không?

Em thì em cũng muốn trau dồi tiếng Anh nhưng chưa có động lực lớn dù em biết đến khi học năm thứ ba cũng cần điều kiện tiếng Anh để có thể đăng ký tín chỉ trong chương trình học chính. Hiện tại em vẫn cảm thấy em chưa muốn học tiếng Anh, em chỉ đang cố gắng trau dồi liên quan nghe nói phát âm tiếng Anh.

Cô được biết điểm TOEIC đầu vào của em là 390 đúng không nhỉ?

Vâng đúng ạ. Em có học TACS2 và xong từ kỳ trước rồi ạ.

Nếu bây giờ có thi lại TOEIC tương tự như bài thi đầu vào thì em dự đoán điểm thi của mình là bao nhiêu?

Em nghĩ điểm số ngang bằng hoặc cao hơn, chắc khoảng 420 ạ.

Theo em lượng từ vựng của em hiện tại là bao nhiêu từ?

Em cũng không rõ, em thấy từ vựng của em hơi ít, em không phản xạ được nhanh, hoặc do em quen với nhiều bạn thành thạo tiếng Anh nên em thấy từ vựng tiếng Anh của em hơi yếu.

Nếu điểm TOEIC là 390 thì theo các nhà nghiên cứu quy đổi sẽ có lượng từ khoảng 1700-2000 từ, em có nghĩ lượng từ của em ở khoảng đó không? Chắc tại em không tự tin vào từ vựng của mình lắm nên nghĩ đến con số bé hơn nhiều ạ.

Bây giờ chúng ta cùng nói về trắc nghiệm thích ứng trên máy tính nhé! Em hiểu gì về trắc nghiệm thích ứng trên máy tính nhỉ?

Em nhớ cô có bảo là từ ngân hàng từ vựng, hệ thống sẽ lựa chọn câu hỏi để kiểm tra từ vựng của em, xem được bao nhiêu lượng từ cơ bản ạ.

Em hiểu thích ứng ở đây là gì?

Là không phải đề cố định, đề linh hoạt và ngắn hơn ạ.

Thích ứng ở đây được hiểu là thích ứng với năng lực của thí sinh. Dựa vào câu trả lời với câu hỏi trước, hệ thống sẽ lựa chọn câu hỏi phù hợp, dễ hơn hoặc khó hơn từ ngân hàng câu hỏi.

Trước đây, em đã có trải nghiệm với trắc nghiệm thích ứng trên máy tính chưa?

Trước khi tham gia bài ấy, em chưa có trải nghiệm nào với trắc nghiệm thích ứng. Em chỉ làm các bài kiểm tra về từ vựng trên các web, và phần mềm TFlat. Mỗi lần em làm với TFlat em có thấy mỗi lượt làm bài có các câu hỏi khác nhau, 100 câu chỉ có vài câu trùng lặp nhưng không chắc đó là trắc nghiệm thích ứng ạ.

Em làm bài kiểm tra thích ứng trên máy tính bằng thiết bị gì?

Em dùng điện thoại.

Theo em việc sử dụng các thiết bị khác nhau có ảnh hưởng đến trải nghiệm của thí sinh không?

Em nghĩ là có, em dùng điện thoại thì đang ngồi chơi cũng có thể mở điện thoại làm và học từ này từ kia, em thấy học như thế khá là thoải mái, còn nếu phải dùng laptop thì phải ổn định tư thế, không thoải mái bằng sử dụng điện thoại.

Bây giờ chúng ta sẽ sang phần chính về các quan điểm của em với các tính năng của trắc nghiệm thích ứng trên máy tính được thể hiện trong bài kiểm tra thích ứng từ vựng tiếng Anh mà em đã thực hiện. Em chia sẻ với cô xem em đồng ý hay phản đối với các nhận định và lý giải cho quan điểm của em.

Thứ nhất, thời gian làm bài kiểm tra thích ứng trên máy tính có ngắn hơn so với bài thi trắc nghiệm hoàn chỉnh 100 câu mà em làm trước đó?

Em nghĩ trải nghiệm của em sẽ ảnh hưởng bởi điều kiện đường truyền của web ạ. Lần trước em làm đề 100 câu trên Google Forms thì chọn đáp án rất nhanh, không bị lag như khi làm trên web với trắc nghiệm thích ứng, ví dụ sau khi em làm xong 1 câu thì chọn đáp án và nhấn NEXT sang câu tiếp theo. Nói chung, không biết do mạng của em hay như thế nào mà em chọn đáp án thì lần nào em cũng thấy phản hồi trên web rất là chậm, như thế nên khi làm em có cảm giác khá khó chịu, thời gian làm còn lâu hơn cả 100 câu kia.

Lúc ấy em dùng mạng ở trường hay 3G nhỉ?

Em dùng mạng ở nhà ạ. Ví dụ như khi em nhấn chọn đáp án A, hệ thống mất khá nhiều thời gian để nhận đáp án đó và cho em câu hỏi tiếp theo. Việc này diễn ra với toàn bộ bài làm chứ không phải 1-2 câu ạ.

Thứ hai là về thao tác thực hiện với trắc nghiệm thích ứng trên máy tính. Em có khúc mắc gì với thao tác khi thực hiện bài kiểm tra thích ứng trên máy tính không?

Với việc lựa chọn đáp án thôi thì khá dễ dàng, giống như khi em thi TOEIC đầu vào, làm xong câu này mới được làm câu tiếp theo. Nhìn chung, thao tác dễ, chỉ cần dùng con trỏ chuột là dùng được ạ.

Thứ ba là bài kiểm tra thích ứng trên máy tính, theo em, có đánh giá chính xác năng lực của em không, trong trường hợp này là năng lực từ vựng tiếng Anh?

Cái này em cũng không rõ năng lực của bản thân em, nhưng em nghĩ là nó đúng, số câu trả lời đúng của em ở khoảng 60-80% thì em nghĩ là nó đúng với khả năng của em ạ.

Theo em các câu hỏi trong bài kiểm tra thích ứng có phù hợp với năng lực của thí sinh không?

Em nghĩ là nó phù hợp với em, vừa có từ em biết, vừa có các từ liên quan đến từ em biết để lựa chọn đáp án. Khi có sự liên quan đó, em nghĩ là đến khi em muốn học lại từ đó, thì em có mối liên hệ với việc đã làm bài kiểm tra rồi.

Khi làm bài kiểm tra thích ứng trên máy tính em có thấy căng thẳng hay khó chịu gì không?

Có, như em chia sẻ, liên quan về mạng và hệ thống. Mỗi lần làm em lại phải đăng nhập lại nên em khá ngại và không thoải mái lắm ạ.

Em có nghĩ trắc nghiệm thích ứng trên máy tính có thể hạn chế gian lận thi cử không?

Em nghĩ là có, đi thi thì một số bạn cố nhớ đáp án, nhưng khi làm trắc nghiệm thích ứng này thì nó sẽ cho câu hỏi mới lạ. Những bạn nhớ đáp án cũng không thể làm được nên chắc sẽ khó chịu vì như bọn em nói thì đề rất là lạ. Theo em nghĩ thì gian lận này không nên vì thi là để phản ánh năng lực của đến đâu để còn tiến bộ ạ. Vì vậy em rất ủng hộ đề thích ứng này.

Sau khi em thực hiện bài kiểm tra thích ứng trên máy tính thì em có được nhận kết quả và các gợi ý về học tập ngay lập tức không? Em thấy thế nào? Có muốn thay đổi gì không?

Em nghĩ là có, nó là một cảm giác gì đó, giống như kiểu là khi em làm tốt bài đấy thì sẽ nhận được lời khen, vì vậy nhận được thông báo kết quả thì sẽ tốt hơn, thích hơn, em nghĩ cái này liên quan đến tâm lý nhiều hơn khi nhận được kết quả và lời khen.

Theo em thì trắc nghiệm thích ứng trên máy tính có phải làm một công cụ để em tự đánh giá hiệu quả không?

Em nghĩ là có. Nếu mà nghĩ đến mảng từ vựng thì em nghĩ nó rất là ổn, nếu là viết thì chắc khó hơn.

Em có thấy trắc nghiệm thích ứng trên máy tính có thể được dùng là một công cụ học tập và ôn luyện thường xuyên không?

Về mảng từ vựng, em nghĩ nó là một công cụ học tập rất hữu ích, ví dụ có thể phát triển theo cách là theo dõi quá trình học được những từ nào rồi, sau đó tạo lập một sơ đồ cây, cho lên đó những từ vựng em sẽ học, liên quan đến và có thể đạt được, kiểu như thế khi em làm xong trắc nghiệm em có thể học những từ liên quan và giúp ích nhiều hơn, theo ý kiến em là như vậy.

Bây giờ sẽ đến phần cuối cùng, em chia sẻ những cảm tưởng hay ấn tượng chung nhất của em về trắc nghiệm thích ứng trên máy tính nhé! Điều gì khiến em hài lòng nhất về trắc nghiệm thích ứng trên máy tính?

Em nghĩ là nó giúp em biết mình đang hỏng ở những kiến thức nào, ví dụ như là khi bài kiểm tra từ những câu hỏi dễ đến những câu hỏi khó, khi em làm đến câu hỏi ở những cái mức khó gần nhau, khi em làm bài và làm đến những câu ấy thì em biết từ vựng của mình đã phát triển đến đâu, và chưa phát triển cái gì. Em nghĩ đó là ưu điểm của nó.

Có điều gì khiến em không hài lòng và muốn thay đổi hay đề xuất gì không?

Em nghĩ là không, với những câu hỏi em làm thì em thấy ổn. Chứ cái đề hoàn chỉnh 100 câu thì dài quá ạ. Em mong chờ trong thời gian tới bài thích ứng có thể hướng tới nhận diện các thì trong tiếng Anh hoặc như kiểu chọn từ đúng trong một bài viết, em nghĩ là cái đó khá là thích hợp với trắc nghiệm thích ứng. Nếu như viết cả câu thì khó dùng trắc nghiệm thích ứng, chỉ ngữ pháp và từ vựng là phù hợp ạ.

Chúng ta đã hoàn thành các nội dung phỏng vấn. Chân thành cảm ơn em đã chia sẻ rất nhiệt tình.