論　文

大規模リスニングテストにおける「妥当性」と「均一性」

－ICプレーヤー試聴体験に参加した高校生の意見分析－

西郡　大[1)*]，倉元直樹[1)]

1）東北大学大学院教育情報学教育部　東北大学高等教育開発推進センター

1．問題と目的

毎年，50万人以上が受験する大学入試センター試験（2007年度の受験者は約51万人）であるが，中でも，最も多くの受験生が受験する科目が「英語」である.「英語」には，平成18（2006）年度より，リスニングテストが導入されており，英語受験者のほぼ全員が受験することから，受験者全員が一斉に実施するリスニングテストとしては，わが国において最大規模のものであると思われる.

リスニングテストのセンター試験への導入について，文部科学省高等教育局学生課大学入試室の報告によれば次のような経緯だったとされる[1)]．平成12年の「大学入試の改善について」（大学審議会答申）[2)] でリスニングテスト導入を図ることの必要性が提言され，文部科学省では，高等学校関係者及び大学関係者等で構成する「大学入学者選抜方法の改善に関する協議」（いわゆる「協議の場」）において検討が進められてきた．一方，平成15年３月には，我が国の英語教育を抜本的に改善するために国として取り組むべき総合的かつ具体的な施策として，文部科学省が先導した「英語が使える日本人の育成のための行動計画」が取りまとめられ，平成18（2006）年度からのセンター試験でのリスニングテスト導入が施策の１つに盛り込まれた．そして，平成15年６月に大学入試センターの「平成18年度からの大学入試センター試験の出題教科・科目について（最終まとめ）」[3)] において，平成18年度から教科「英語」でリスニングテストを実施することが明記された．その後，大学入試センターや文部科学省の「協議の場」において最終的な検討を経たのち，平成15年11月５日に公表された.

ところで，センター試験にリスニングテストが導入される趣旨とは何なのであろうか．現行の高等学校学習指導要領における外国語の目標には，「外国語を通じて，言語や文化に対する理解を深め，積極的にコミュニケーションを図ろうとする態度の育成を図り，情報や相手の意向などを理解したり自分の考えなどを表現したりする実践的コミュニケーション能力を養う」とあり，「聞くこと」，「話すこと」，「読むこと」，「書くこと」の４領域を含むコミュニケーション能力が重要視されている[4)]．こうした能力の１つである「聞くこと」の評価には，「“聞くこと”の達成度を実際に音声を使って評価するために，リスニングテストを実施することが望ましい」とされる[5)]．さらに，リスニングテスト導入は，下級学校における英語の指導方法の改善，モティベーションや学習意欲に極めて大きな影響を与えるものと考えられており，特に，毎年約50万人以上の受験生が受験するセンター試験の位置づけと規模を考えれば，リスニングテスト導入によって得られる効果は相当に大きいものだと期待されているのである[6)].

かくしてセンター試験にリスニングテストが導入されたわけであるが，導入過程おいて最も問題視されてきたことが試験の公平性に関する問題である．この問題は，「大学入試センター試験へのリスニングテストの導入が要望されながら，見送られてきたのは，すべての受験生に対して音が聞こえる環境等を絶対的に均一にすることが不可能と考えられてきた」という大学

*）連絡先：980-8576　宮城県仙台市青葉区川内28　東北大学高等教育開発推進センター高等教育開発部入試開発室

審議会答申の総括[7]が最も端的に示している．つまり，50万人以上の受験生が全国一斉に，それも同一時間帯にリスニングテストを受験するためには，公平性の確保が非常に難しい問題として立ちはだかってきたのである．

こうした公平性に関して，導入決定以前は，スピーカーを用いたリスニング方式を中心に何度か検討されてきた．例えば，昭和57年に，大学入試センターの聴解試験プロジェクトチームが行った試験室内の音響条件を均等にするスピーカーの配置を検討したものや[8]，同じく大学入試センターの研究開発部を中心とした研究グループの受験生の着席位置によるスピーカーからの音声の大きさの違いが解答に及ぼす影響を定量的に評価したもの[9]が挙げられる．両研究の結果とも，全受験生の試験環境を同条件に統一するのは難しいというものであった．しかしながら，リスニングテスト導入への志向はコミュニケーション能力重視が強まるとともに縮減することなく，むしろ期待されるものとなり，結果的に実現されるに至った．こうした議論からは，「聞く力」を測定するために「妥当」だと考えられるテスト実施，いわゆる「妥当性」[10]の問題と全受験生の試験実施環境が同一条件，公平でなければならないという「均一性」の問題という対立構造を見出すことができよう．

リスニングテスト導入決定後，ICプレーヤーを用いた個別音源方式の試験実施を方針としたことで，リスニング試行テストやモニター試験等を通して実施方法の改善のための情報収集やICプレーヤーの品質，信頼性向上のための試作機を用いた実地検証等が入念に行われ，公平性確保に向けた十分な検証がなされてきた[11]．その結果，過去２回のテストは，全体の規模から考えれば，極めて少数のトラブルに留まり，センター試験終了後，ICプレーヤーのトラブルによる「再開テスト」の人数やその試験実施条件における公平性に関して報道はされるものの，リスニングテストの存在自体を根本から揺るがすような大きな社会問題にはなっていないように思われる．これは，社会全体からみれば，コミュニケーション能力として必要な「聞く力」を評価するためのリスニングテストの必要性が，ある程度是認されているのではないかとみることができる．

そこで本研究では，リスニングテスト導入３年目を迎える時期において，センター試験を受験すると考えられる高校生を対象に，センター試験で用いるものと同じICプレーヤーを用いて，大ホールにおいて一斉にリスニングテストの体験を行った．試聴体験実施後，自分が受験するであろう大学入試センターのような大規模かつ「ハイステークス」な試験でのリスニングテストに対する意識を調査した．特に，リスニングテストをめぐる「妥当性」と「均一性」という観点を中心に据え，どのような意識が潜在しているかを検討することが主な目的である．同時に，彼らがどのようなリスニングテストの実施様式を選好するのかについても検討した．

２．大学入試センター試験のリスニングテスト[12]

２．１．実施概要

外国語「英語」の受験者は，従来の筆記試験とリスニングテストの両方を受験しなければならない．筆記試験終了後40分の休憩を挟み試験が開始される．試験時間は60分であり，ICプレーヤー等の配布や音声認識が行われた後，30分で音声問題を解答することになっている．ICプレーヤーの主な機能は，音量調整と再生のみである．巻き戻して問題を何度も聞き直すことはできない．

２．２．再開テスト[13]

解答時間中に，ICプレーヤーの不具合があった場合や，問題冊子に印刷不鮮明，ページの乱丁・落丁があった場合，不慮の事故等により続行することが出来ない場合など，監督者の指示で試験が中断される．この場合，リスニングテスト終了後，別のICプレーヤーにより当初解答していたものと同じ試験問題を使って，中断箇所以降のみについてテストが再開できる．不具合のあったICプレーヤーは調査のため回収され，調査の結果，虚偽の申し出であることが判明した場合は不正行為として処罰される．実施された過去２年間の再開テスト受験者は，平成18年度が457人（全体の約0.09%），平成19年度が381人（全体の約0.07%）であった[14]．

3．方法

3．1．調査概要

東北大学高等教育開発推進センターが，ある県立高校（東北地方）と協力して企画したシンポジウム[15]を利用して調査を実施した．同シンポジウムは，主に協力校の高校生，教師，保護者，一般者を対象として「テスト」について考えるものであり，筆者らを含め３人の講演者が，「センター試験リスニングテスト体験 －テストの理論と技術－」，「『公平な入試』は実現可能か？ －社会心理学的アプローチ－」，「テストと進路選択 －国家公務員採用試験におけるテスト－」というタイトルで30～40分程度講演した．同シンポジウムを利用したのは次の理由からである．まず，本企画が「テスト」を問題にしたものであること．そして，大学入試センターの協力を得て，本試験で用いるものと同じICプレーヤーを900個借りることができたこと．さらに，１つの会場において数百人規模で試聴体験が一斉にできるということである．

3．2．調査手続き

【対象者】シンポジウムの参加者全員である．その内訳は，１年生と２年生の生徒約630人．引率の教師10名弱．保護者・一般参加者30名程度である[16]．協力高校は，多くの生徒が大学受験をする東北地方における共学の進学校である．なお，生徒は「総合的な学習」の一環として参加しており，引率として各クラスの担任が同行している．保護者・一般参加者は，高校からの周知や東北大学高等教育開発推進センターのホームページなどから募って参加してきた希望者である．

【会場】同シンポジウムは，収容人数が最大で約1500名の大ホールを利用した．本会場は３階席まであるが，その１階席（930席）のみを使用した．通常は，コンサートや講演などに利用されるホールである．

【質問紙の配布】生徒が着席するスペースは，予め指定されていたので，生徒たちが着席する前に，指定された座席に封筒を配布した．封筒には，プログラム，講演資料等，シンポジウムのアンケート，試聴用のICプレーヤーに加え，質問紙を同封した．

【質問紙の内容】質問紙の内容は，「Ｑ１．回答者のプロフィール（３項目）」「Ｑ２．各種経験の有無（３項目）」「Ｑ３．高校で学ぶ英語に対する認識（５項目）」「Ｑ４．リスニング試験一般に対する認識（７項目）」「Ｑ５．リスニング試験でのトラブルに対する認識（３項目）」「Ｑ６．試聴体験を通じての感想（５項目）」「Ｑ７．『妥当性』と『均一性』について（強制選択法）」「Ｑ８．自分が好むリスニング試験の実施様式（単一回答法）」「Ｑ９．リスニング試験に対する感想（自由記述法）」から構成される（巻末資料）．Ｑ３～Ｑ６までの各項目は，「そう思わない（１点）」「あまりそう思わない（２点）」「どちらとも言えない（３点）」「少しそう思う（４点）」「そう思う（５点）」の５件法で求め，（　）内に示した得点を付与した．なお，Ｑ３の最後の項目は，自由記述形式である．

【回答のタイミング】質問紙への回答の前提は，以下の２つの理由から最初の講演終了後である必要がある．まず，リスニングテスト体験後でなければならないこと．そして，「テストとは何か」ということに関連して，リスニングテストの妥当性ということを理解してもらうために，測定技術としての「テスト」という観点から，テストの「妥当性」や「信頼性」の概要について説明する必要があるためである．そこで，最初の講演者（この講演者は，教育測定，特に，テストの性質や仕組みにおける技術や理論，いわゆる「テスト理論」に関して専門的な知識を有する研究者である）に，テストの「妥当性」を中心に説明してもらった[17]．したがって，回答のタイミングに際しては，２人目の講演終了後の休憩時間，もしくは，シンポジウムの全プログラム終了後に質問紙へ回答するようにアナウンスした．なお，１人目の講演終了から休憩までの時間は約30分であり，休憩時間は20分間である．

【質問紙の回収】会場の出入り口付近に回収箱とスタッフを配備し，休憩時間以後からプログラム終了まで継続的に回収した．なお，シンポジウムに対するアンケートも同時に回収している．

3．3．分析方針

分析で使用するデータは，生徒のみの回答とする．理由は，教師，保護者，一般参加者のサンプルサイズが小さいこと．そして，分析の対象として，実際にセンター試験を受験する生徒を想定することが妥当だと

考えられるためである．分析の中心に据える柱は,「妥当性が高いテスト」,「均一性が高いテスト」どちらが公平・公正だと思うかという強制選択法（Ｑ７）と，大規模リスニング試験に期待する実施様式に対する単一回答法（Ｑ８）である．まず,生徒の性別,学年によってＱ７，Ｑ８に対する回答に違いが生じるかを検討した．次に，Ｑ３～Ｑ６の回答から特徴的な要因を抽出し，これらの要因とＱ７，Ｑ８との回答群，その他の属性も考慮に入れて探索的に比較検討した．なお，質問紙ではセンター試験を含む一般的な大規模リスニングテストを想定して「リスニング試験」という用語を用いたが，回答者はセンター試験のリスニングテストをイメージして回答していると考えられるために，本稿では，質問紙で表示したもの以外は,「リスニングテスト」と統一して表記した．

４．結果

４．１．回答者のプロフィール

回収された質問紙は555件であった．その中から，無回答項目が多いもの，明らかに真面目に回答していないもの32件を除く523件を分析対象とした．属性の内訳は,生徒490人（１年生:258人<男子:125人,女子:133人>，２年生：227人<男子：98人，女子：129人>，無回答５人），教師８人（男性：６人，女性：２人），保護者・一般23人（男性：４人，女性：19人），無回答２人であった．

センター試験を受験しようと考えている生徒は450名（91．８％），受験を考えていない生徒は８名（1.6%），無回答32名（6.6%）であった．

生徒に限定した海外（英語圏）在住の経験者は６名（1.2%），学校以外の英会話教室や海外留学で英会話を学んだことある者は103名（21%），模擬試験などでICプレーヤーを用いたリスニングテストの経験者は27名（5.5%）であった．

４．２．Ｑ７とＱ８の集計とその関係

Ｑ７の質問内容は,「次の２つのうち，どちらが公平・公正なテストだと思いますか．１つだけ数字に○をつけて下さい」というものであり，各項目に対する回答者は「Ｉ．試験内容の妥当性は高いが，受験生全員の試験実施条件を完全に均一にするのが難しいテスト」が299名（61%),「Ⅱ．試験の実施条件は完全に均一性を確保できるが，試験の内容的な妥当性が低いと思われるテスト」が184名（37.6%），無回答が７名（1.4%）であった．

Ｑ８の質問内容は,「受験者にとって重要であるセンター試験のような大人数が受験するリスニング試験の実施様式の中で，次のうちから最も適当だと思うもの１つに○をつけてください」というものであり，各項目に対する回答者は,「Ｉ．個別音源方式（ICプレーヤー）による試験」が239名(48.8%),「Ⅱ．教室のスピーカーによる一斉放送試験」が131名（26.7%),「Ⅲ．筆記によるアクセント・発音の問題による代用」が55名(11.2%),「Ⅳ．リスニングに関する試験は必要ない」が62名（12.7%），無回答が３名（0.6%）であった．

まず，Ｑ７，Ｑ８に対する回答が「性別」と「学年」によって影響を受けるかついて，西郡・木村・倉元（2007）[18] に倣って対数線形モデルを用いて検討した結果,特徴的な影響は見出されなかった．したがって，Ｑ７とＱ８のクロス表に関する分析は,「性別」,「学年」に分けることなく「生徒」という１つのカテゴリーで分析を行うことにする．

次に，西郡・木村・倉元（前掲書）に倣いＱ７とＱ８の関係を検討した．全体の傾向を見るために，カイ２乗検定により各変数間の連関をみたところ，５％水準で有意差が確認された．また，Ｑ７とＱ８の各項目間の関係をみるために残差分析を行い,「調整された残差（adjusted residual)」を算出した（表１）.「調整された残差」には標準正規分布が仮定されており，±1.96以上の値を示せば，５％水準で有意差があるとされる（有意であるものは,表１の「調整された残差」の値をボールドで示した）．その結果,「Ｑ８_Ｉ．個別音源方式（ICプレーヤー）による試験」と「Ｑ８_Ⅲ．筆記によるアクセント・発音の問題による代用」において，Ｑ７_Ｉ，Ｑ７_Ⅱの選択者群に有意差がみられた．なお，Ｑ７_ＩとＱ７_Ⅱのオッズ（Ｑ７_ＩとＱ７_Ⅱの件数の比）からは，Ｑ８のほとんどの選択項目において，Ｑ７_Ｉの回答数が全体的に多いことが示されている．

表１．Ｑ７とＱ８の関係

様式	Ｑ７＿Ⅰ	Ｑ７＿Ⅱ	オッズ	合計
Ｑ８＿Ⅰ	159 **(2.0)**	80 **(-2.0)**	1.98	239
Ｑ８＿Ⅱ	72 (-1.6)	56 (1.6)	1.29	128
Ｑ８＿Ⅲ	26 **(-2.2)**	28 **(2.2)**	0.93	54
Ｑ８＿Ⅳ	42 (1.2)	19 (-1.2)	2.21	61
合計	299	183	1.63	482

（　）内は，「調整された残差」

４．３　試聴体験後のリスニングテストに対する認識の因子構造

各変数の構造を明らかにするために，Ｑ３～Ｑ６の項目について因子分析を行った．これらの項目分析の結果は，全項目間において r =.70以上の相関が認められず，度数分布の確認においてはＱ５－２のみに偏りがみられた．そのためＱ５－２を除いた19項目による探索的因子分析を行った．その結果，スクリープロットより３因子に固定し，算出された因子負荷量の.30未満の項目を除いて再度，因子分析（主因子法・バリマックス回転）を行った（表２）．

第Ⅰ因子は，「『英語』は好きだ」，「英語教育は，『聞く』，『話す』などのコミュニケーションを重視したものがよい」などの６項目が高い負荷量を示した．「英語」や「コミュニケーション」を積極的に受け入れているものであると考えられるため「英語･コミュニケーション積極的受容」と命名した．

第Ⅱ因子は，「操作に気をとられて，試験に集中で

表２．Ｑ３～Ｑ６の因子分析から得られた因子パターン（バリマックス回転後）

	Ⅰ	Ⅱ	Ⅲ	共通性
Ⅰ．英語・コミュニケーション積極的受容（α=.68）				
Ｑ３-2.「英語」は好きだ	.71	.02	.12	.51
Ｑ３-3. 英語教育は，「聞く」「話す」などのコミュニケーションを重視したものがよい	.59	.03	.10	.36
Ｑ３-1.「英語」は得意だ	.55	.00	-.02	.30
Ｑ３-4. 入試での英語もコミュニケーションを重視した試験がよい	.47	.00	-.00	.22
Ｑ４-6. リスニング試験がなければ，リスニングの勉強に興味が無い	-.44	.08	-.10	.21
Ｑ４-7. リスニング試験は，自分にとって有利だと思う	.33	-.06	.13	.13
Ⅱ．操作・条件に対する不安（α=.65）				
Ｑ６-3. 操作に気をとられて，試験に集中できないかもしれない	.06	.77	.02	.60
Ｑ６-1. 一度も操作の練習をせずに，本試験を受けるのは不利だ	-.00	.53	.03	.28
Ｑ６-2. 操作においては，何も問題なく上手に使いこなせる自信がある	-.06	-.53	-.02	.28
Ｑ６-4. 周囲の雑音が気になって，試験に集中できないかもしれない	-.04	.50	-.08	.26
Ｑ６-5. 周囲の受験者もほとんど自分と同じように聞こえていると思う	.08	-.42	.16	.21
Ｑ５-3. 故障したICプレーヤーに当たり，再開テストを行ったとしても不利に感じる	-.06	.30	-.07	.10
Ⅲ．リスニングテストへの肯定的信念（α=.68）				
Ｑ４-2.「聞く力」を身につける上で，リスニング試験の勉強は役に立つと思う	.06	-.05	.78	.62
Ｑ４-1. リスニング試験は，「聞く力」を測定するために妥当な試験だと思う	.09	-.10	.73	.55
Ｑ４-4. 日ごろの努力次第で，リスニング試験の得点は上がると思う	.21	-.06	.47	.27
因子寄与	2.1	1.7	1.1	

きないかもしれない」,「周囲の雑音が気になって，試験に集中できないかもしれない」などの６項目が高い負荷量を示した．ICプレーヤーの操作や試験の実施条件に対する不安を表したものであると考えられるため「操作・条件に対する不安」と命名した．

第Ⅲ因子は,「『聞く力』を身につける上で，リスニング試験の勉強は役に立つと思う」,「リスニング試験は,『聞く力』を測定するために妥当な試験だと思う」など３項目が高い負荷量を示していた．リスニングテストに対して肯定的に捉えていることから「リスニングテストへの肯定的信念」と命名した．

なお，Cronbachのα係数は各因子とも．65程度であり，ある程度の内的一貫性は保たれていると考えられる．

４．４　各因子の尺度得点からみる性別，学年による差

因子分析で抽出された３種類の尺度得点の平均得点を，性別と学年を要因とする二元配置の分散分析を行った．その結果,「操作・条件に対する不安」のみ，性別のみに主効果がみられ（F（1, 448）=10. 73, p<. 01），女子の得点が高いことが示された（表３）．

表３．性別，学年による各因子の平均得点の比較

	男子（n=210）		女子（n=242）		主効果		交互作用
	１年(n=120)	２年(n=90)	１年(n=123)	２年(n=119)	性別	学年	
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*F*値	*F*値	*F*値
英語・コミュニケーション積極的受容	2. 79（. 70）	2. 82（. 69）	2. 93（. 65）	2. 82（. 71）	1. 02	0. 40	1. 31
操作・条件に対する不安	2. 82（. 74）	2. 73（. 82）	2. 96（. 64）	3. 04（. 74）	10. 73**	0. 01	1. 54
リスニングテストへの肯定的信念	3. 51（. 86）	3. 58（. 93）	3. 72（. 81）	3. 61（. 86）	2. 20	0. 04	1. 27

$^{**}p$ < . 01

４．５．公平･公正なテストとしての「妥当性」「均一性」からみる各尺度得点の比較（Ｑ７の分析）

まず，Ｑ７_ⅠとＱ７_Ⅱの選択者群に分け，各因子における尺度得点の平均点の差についてt検定を行ったところ，有意差はみられなかった（表４）．また，判別分析を用いてＱ７に対する回答を予測することで各因子の影響力の有無を調べてみたが，固有値が低く線形判別関数によってうまく判別されていないために十分に予測できないものであった．そのため，各因子の尺度得点とＱ７の回答群間に有意な関係はないと考えられる．

表４．Ｑ７の回答群による各因子の平均得点の比較

	Q7-Ⅰ（n=278）		Q7-Ⅱ（n=174）		t値
	M	*SD*	*M*	*SD*	
英語・コミュニケーション積極的受容	2. 89	. 70	2. 77	. 68	1. 84
操作・条件に対する不安	2. 87	. 76	2. 93	. 71	. 83
リスニングテストへの肯定的信念	3. 59	. 88	3. 64	. 83	. 61

次に，Ｑ７の回答群と「Ｑ２-2．学校以外の英会話教室や海外留学などで英会話を勉強したことがある」という海外留学や英会話の経験（以降,「海外英会話経験」と表記）を要因とする二元配置の分散分析を行った（表５）．その結果,「英語・コミュニケーション積極的受容」において，Ｑ７と海外英会話経験に有意な主効果がみられた（F（1, 444）=5. 19, p<. 05：F（1, 444）=18. 8, p<. 001）．また,「リスニングテストのへの肯定信念」では，交互作用（F（1, 444）=3. 90, p<. 05）が確認され, 海外留学･英会話の経験者，かつ，妥当なテストを公平・公正だと考える者は，リスニングテストに対して強い肯定観を抱いている傾向が示された．

表５．Ｑ７と海外英会話経験による各因子の平均得点の比較

	Q7_Ⅰ（n=276）		Q7_Ⅱ（n=172）		主効果		交互作用
	有り（n=57）	無し（n=219）	有り（n=36）	無し（n=136）	Q7	海外英会話経験	
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*F*値	*F*値	*F*値
英語・コミュニケーション積極的受容	3.25（.74）	2.80（.66）	2.97（.74）	2.71（.65）	5.19*	18.8***	1.44
操作・条件に対する不安	2.85（.79）	2.88（.75）	3.17（.72）	2.87（.70）	3.18	2.34	3.63
リスニングテストへの肯定的信念	3.84（1.0）	3.52（.83）	3.56（.89）	3.65（.81）	.540	1.37	3.90*

*p<.05　***p<.001

４．６．リスニングテスト実施様式選好別にみる各尺度得点の比較（Ｑ８の分析）

センター試験のように重要で大規模なリスニングテストにおいて，受験者はどのような実施様式を期待するのかについて，Ｑ８を要因として４水準（Ⅰ～Ⅳの選択群）一元配置の分散分析を行った（表６）．有意な主効果が確認された場合にはTukey法にて多重比較を行い，２群間に有意差（５％水準）がみられたものは不等号（<），有意差がみられなかったものは（=）を用いて表記した．

表６．期待する実施様式別の各因子の平均得点の比較

	Q8					多重比較（Tukey法）
	Ⅰ（n=226）	Ⅱ（n=119）	Ⅲ（n=52）	Ⅳ（n=58）	主効果	
	M（SD）	M（SD）	M（SD）	M（SD）	F値	
英語・コミュニケーション積極的受容	2.94（.69）	2.86（.66）	2.66（.58）	2.59（.77）	5.29**	Ⅰ＞Ⅲ，Ⅰ＞Ⅳ
操作・条件に対する不安	2.75（.69）	3.02（.78）	3.04（.73）	3.08（.76）	6.04***	Ⅰ＜Ⅱ＝Ⅲ＝Ⅳ
リスニングテストへの肯定的信念	3.83（.78）	3.54（.82）	3.40（.80）	3.06（.99）	15.06***	Ⅰ＞Ⅱ＝Ⅲ＝Ⅳ，Ⅱ＞Ⅳ

**p<.01　***p<.001
多重比較の有意水準は５％

その結果，３因子とも主効果（F（3, 451）=5.29，p<.01：F（3, 451）=6.04，p<.001：F（3, 451）=15.06, p<. 001）が確認された．多重比較の結果,「英語・コミュニケーション積極的受容」において，「Ⅰ．個別音源方式（ICプレーヤー）による試験」が「Ⅲ．筆記によるアクセント･発音の問題による代用」と「Ⅳ．リスニングに関する試験は必要ない」に比べて有意に高いことが示された．また,「操作･条件に対する不安」では，「Ⅰ．個別音源方式（ICプレーヤー）による試験」がその他の実施様式より有意に低かった．さらに，「リスニングテストへの肯定信念」では，「Ⅰ．個別音源方式（ICプレーヤー）による試験」がその他の実施様式より有意に高く，かつ,「Ⅱ．教室のスピーカーによる一斉放送試験」が「Ⅳ．リスニングに関する試験は必要ない」よりも有意に高いことが示された．なお，多重比較においては，群ごとのサンプルサイズが異なるために調和平均が用いられている．

また，Ｑ７を要因に含めた二元配置の分散分析を行ったところ，同じく，Ｑ８において主効果（F（3, 443）=5.05，p<.01：F（3, 443）=6.89，p<.001：F（3, 443）=12.19，p<.001）が確認され，多重比較の結果も上記の傾向と全く同じであった．

次に，Ｑ８の回答群と海外英会話経験を要因とする二元配置の分散分析を行った（表７）．その結果，３因子すべてにおいてＱ８に主効果がみられた（F（3, 443）=5.15，p<.01：F（3, 443）=4.83，p<.05：F（3, 43）=15.73, p<.001）．「英語･コミュニケーション積極的受容」では，海外英会話経験における主効果（F（1, 443）=4.87，p<.05），さらに交互作用も示された（F（3, 443）=2.85，p<.05）．一方，「リスニングテストへの肯定的信念」においても交互作用が確認された（F（3, 443）=2.71，p<.05）．なお，Ｑ８の多重比較は，表６と同じ結果が示された．

表７．Q8と海外英会話経験による各因子の平均得点の比較

	Q8-Ⅰ（n=226）		Q8-Ⅱ（n=117）		Q8-Ⅲ（n=51）		Q8-Ⅳ（n=57）		主効果		交互作用
	有り（n=55）	無し（n=171）	有り（n=24）	無し（n=93）	有り（n=8）	無し（n=43）	有り（n=6）	無し（n=51）	Q8	海外英会話経験	
	M(SD)	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*	*M(SD)*	*F*値	*F*値	*F*値
英語・コミュニケーション積極的受容	3.26 (.75)	2.83 (.64)	3.13 (.71)	2.79 (.64)	2.35 (.57)	2.71 (.57)	3.06 (.70)	2.53 (.77)	5.15**	4.87*	2.85*
操作・条件に対する不安	2.86 (.71)	2.72 (.68)	3.08 (.84)	2.99 (.77)	3.00 (.88)	3.07 (.71)	3.61 (.91)	3.03 (.74)	4.83**	2.57	.86
リスニングテストへの肯定的信念	4.04 (.74)	3.76 (.79)	3.53 (1.04)	3.53 (.75)	3.29 (1.00)	3.40 (.77)	2.39 (1.14)	3.10 (.93)	15.73***	1.10	2.71*

*p<.05　**p<.01　***p<.001

５．考察

Ｑ７とＱ８の分析結果より，妥当性が高いテストを公平・公正だと考える者の方が，現在のセンター試験におけるリスニングテストの実施形態を支持する傾向があるものの，均一性が高いテストを公平・公正だと捉える者にも，「個別音源本式（ICプレーヤー）」を期待する者が決して少なくない．つまり，「ICプレーヤーを使った試験は，席によってムラがないからいいと思う」という回答者の意見が示すように，スピーカーによる音の聞こえ方の不均一性を解消するために導入された個別音源方式の狙いが反映された結果だと言える．

一方，均一性が高いテストを公平・公正だと捉える者には，「Ｑ８－Ⅱ．教室のスピーカーによる一斉放送試験」を実施様式として期待する者が相対的に多い．スピーカーによるリスニングテストの不均一性を解消するために個別音源方式が導入されたという経緯を考えれば，こうした傾向は矛盾しているように思える．この矛盾を解釈するための主要な要因として考えられるのが，因子の１つとして抽出された「操作・条件に対する不安」である．具体的には，「１人1つだと自分１人のような気がして緊張してしまうけど，教室のスピーカーなら皆同じものを聞いていると思えるし，いつも通りに出来ると思う」という自分が耳にする音源だけは間違いなく同室で受験する他の受験生と同じであるという安心感や安定感を志向する意見や「普段の学校で行う試験では，教室のスピーカーによる一斉放送なので，ICプレーヤーは精神的に不安が伴う気がします」という普段の試験形態からの乖離に不安を感じている意見がみられる．

こうしたリスニングテストに対する不安に関する意見は，「テスト不安」からの解釈が可能であろう．Mandler&Sarason[19] によると「テスト不安」とは，「テストでよい成績を取るのに必要な反応を妨害する不適切な反応」とされる[20]．つまり，リスニングテストというこれまでは一般的でなかった試験の導入に伴い試験で使用する機器や条件に対してストレスを感じ，本来の能力を万全に発揮できないことへの不安だと解釈できる．特に，テスト不安には性差がみられ，女性においてテスト不安が一般的に高いことが古くから知られている[21]．本研究の「操作・条件に対する不安」の尺度得点を性別に比較した結果からも女子の平均点が有意に高いことが示されている．Spielberger[22] によれば，「テスト不安」には，不安検査で測定されるような個人の心理特性である「特性不安（trait anxiety）」と場面に応じて常に変化する「状態不安（state anxiety）」があるとされるが[23]，機器故障や操作条件などが関わるＩＣプレーヤーを用いたリスニングテストに対する不安は，後者の不安に位置づけられるだろう．Ｑ７_Ｉの選択者にＱ８_Ⅱを選択する者が少なくないことを鑑みても，リスニングテスト場面に関わる「テスト不安」は無視できない問題だと考えられる．とすれば，今後，ICプレーヤーによる個別音源方式の実施を安定化させていくためには，機器の品質，実施・運用面はもちろんのこと，高校生の日常の試験環境などとの比較から生じる受験生が抱くであろう「状態不安」に配慮した工夫も必要になるかと思われる．

次に，因子分析によって抽出された３因子の尺度得点をＱ７とＱ８の回答群別にみた比較について考察をする．分析当初，妥当なテストを公平・公正だと考えるＱ７_Ｉの選択群では,「Ｑ４-1．リスニング試験は,『聞く力』を測定するために妥当な試験だと思う」という項目を含む「リスニングテストへの肯定的信念」の平均点が相対的に高くなると予想していたが，それに反して,両群間の平均点に有意差は無かった(表４)．むしろ，Ｑ７__Ⅱ群における平均点の方が有意差こそ示されないが幾分か高い値であった．そのため，平均値を見る限り，両群ともそれなりにリスニングテストに対して肯定的に捉えているとみることができる．おそらく，リスニングテスト自体には肯定的な認識を持つものの，センター試験のような「ハイステークス」な試験においては，妥当性は高いけれども均一性の確保が難しい試験よりも，安定かつ均一性が十分に高い試験を望む動機が作用したことが考えられる．

大規模リスニングテストで期待する実施様式別（Ｑ８）にみると，３因子全ての平均得点に主効果がみられた．まず,「英語・コミュニケーション積極的受容」では,「Ｉ．個別音源方式（ICプレーヤー）による試験」が「Ⅲ．筆記によるアクセント・発音の問題による代用」と「Ⅳ．リスニングに関する試験は必要ない」と比較して平均点が高く,「Ⅱ．教室のスピーカーによる一斉放送試験」とほぼ同得点であることから，リスニングテスト実施自体に肯定的な反応を示している．同様の傾向は,「リスニングテストへの肯定的信念」にも見られ,「Ｉ.個別音源方式(ICプレーヤー)による試験」,「Ⅱ．教室のスピーカーによる一斉放送試験」を期待する者にリスニングテストへの強い肯定観がみられた．

一方,「操作・条件に対する不安」では,「Ｉ．個別音源方式（ICプレーヤー）による試験」が他の３つの実施様式よりも平均点が低い．つまり，ICプレーヤーの選択者には操作・条件に対する不安が少なく，他の実施様式を選択した者は，ICプレーヤー特有の操作や機器故障に対して不信を抱いているために操作・条件に対する不安が高いことを示している．

最後に，海外英会話経験の影響を考察する．Ｑ７とＱ８の各回答群と海外英会話経験の二元配置の分散分析を行った結果,「英語・コミュニケーション積極的受容」のみ，海外英会話経験要因に主効果がみられたことから，海外留学・英会話の未経験者よりも経験者の方が，英語教育や英語でのコミュニケーションを積極的に受容している傾向が示された．これは英会話への距離が近かった分，親近感が生じるために示された傾向であろう．また，同分析では，Ｑ７にも主効果が確認された．しかし，Ｑ７_ＩとＱ７_Ⅱの選択群に分けてのt検定，Ｑ８との二元配置による分散分析において有意差が確認されていないことから，本主効果に関してはデータの微妙なブレによって生じた可能性が否定できない．一方,「リスニングテストへの肯定的信念」では，海外英会話経験要因とＱ７で交互作用が示された．この背景には，自分の経験を活かすためにリスニングテストを活用したいという動機[24]やコミュニケーション能力の１つとして「聞く力」は重要であることを経験的に知るが故に，妥当性の高いテストを公平・公正であると考え，さらにリスニングテストに対して肯定的な態度を示していることが考えられる．

６．結語

今後，リスニングテストが今以上に受容されていくためには何が必要であろうか.リスニングテストが「聞く力」を測定するために実施されたと考えるならば,「リスニングテスト＝『聞く力』の測定 ＋ α」という図式が成り立つと考えてもよいだろう．

例えば，妥当性の高いテストを公平・公正だと考える者は，測りたい能力を測定する妥当性の高いテストとしてリスニングテストを考える故に，当該テストへの肯定観が強いかと思われたものの，妥当性の高いテストと均一性の高いテストのどちらが公平・公正なテストだと思うかということについて強制選択法（Ｑ７）で選ばせた結果には，その傾向は確認されなかった．むしろ，均一性の高いテストを公平・公正だと考える者と比較して大きな違いは見られず，両者ともそれなりにリスニングテストを肯定的に捉えていることが示された．これは，多くの回答者にとってリスニングテストが妥当で重要なテストであると考えられつつも，何が公平・公正であるかという判断の局面に立たされたとき，各人の公平・公正観が異なっているために生

じた結果だといえる．特に，自分の利害が直接的に絡む「ハイステークス」な試験では，他者との条件面での相違が自分に不利に作用することを嫌うがために，個人の公平・公正認識は非常に多元的になることが予想される．加えて，リスニングテストに特徴的なものかもしれないが，「操作・条件に対する不安」のような「テスト不安」が，こうした認識を助長していることも否めないだろう．

つまり，先に示した図式のように，「リスニングテスト」とは，「『聞く力』の測定」というリスニングテストの「妥当性」の部分以外に「操作や条件に対する不安」や「各人が持つ公平・公正観」のような葛藤を示す「α」という諸問題が不可避的に生じるのである．

当然，α部分が小さく，「妥当性」の部分が多くを占めていれば妥当性が高い公正な試験であると認識されるだろうが，反対に，αの部分が大部分を占めるようになれば，リスニングテストは妥当性が低いテストということになり，不公正なテストとして当事者にみなされかねない[25]．となれば，α部分は出来る限り最小化される必要がある．

その具体的な方策の１つに，「操作・条件に対する不安」を解消するための，ICプレーヤーの操作練習や個別音源方式での模擬試験等の徹底が挙げられる．これは，「操作・条件に対する不安」因子に含まれる「Q６-1．一度の操作の練習をせずに，本試験を受けるのは不利だ」という項目の尺度得点が最も高いことからも有効な方策だと思われ，受験生の「テスト不安」の解消，操作や条件面での混乱を最小限に留められることが期待される．ひいては，当事者たちの公平・公正認識にも影響することでα部分の総和を小さくすることができるだろう．

しかしながら，同時に留意すべき点も見えてくる．例えば，ICプレーヤーに１つでも故障があれば，全員が試験をやり直すというような極端に「均一性」を重視した公平性の確保は，テスト実施のコスト増や効率性の低下に繋がりかねない．また，リスニングテスト対策に向けた過度な操作トレーニングや得点獲得のためのリスニングテクニックの習得といった短絡的な方向へ舵が切られるのであれば，コミュニケーション能力としての「聞く力」を評価することを念頭においたリスニングテストに対するパラドックスに陥りかねない．類似した構図が，高校までの学習到達度を測る大学入試への過度の受験対策が，「受験テクニック」などと批判されてきたことに見られるだろう．さらに穿った見方をすれば，「英語教育」にICプレーヤー操作の習熟が必要であるのかという本質的な問題すら生じてしまう．

今後，高等教育のユニバーサル化にしたがい，規模の大小に関わらず，さらなる多様化が入試において促進されることは否定できない．となれば，リスニングテストに限らず，新たなテストを模索するとなったとき，同じような構図の問題が生じることが予想されるだろう．また，ある局面では，現状からは想像し難い「新たなテスト不安」が喚起されるかもしれない．今後の入試策定において，「新たなテスト」の模索が必要となるとき，本研究で得られた知見が活用されることを期待したい．

〔付記〕

本研究は，東北大学高等教育開発推進センター平成19年度高等教育の開発推進に関する調査･研究経費（センター長裁量経費）である「大学入試学（admission studies）構築のための基礎研究　－人材育成の観点から－（研究代表者：倉元直樹）」の成果の一部である．

〔参考文献・註〕

１）文部科学省高等教育局学生課大学入学室．「英語」のリスニングテストの導入について，大学入試フォーラム，No26，24-29，2003.

２）大学審議会『大学入試の改善について（答申）』，2000.

３）大学入試センター事業部．平成18年度からの大学入試センター試験の出題教科・科目等について，大学入試フォーラム，No26，５-12，2003.

４）旧学習指導要領ともコミュニケーションを重視するという点においては大きな差異はない．

５）大学入試センター．平成９年度からの大学入試センター試験の出題教科･科目等について－中間まとめ－，大学入試フォーラム，No17，125-132，1994.

６）文部科学省高等教育局学生課大学入学室，前掲書．

7）大学審議会，前掲書.

8）聴解試験プロジェクトチーム.『共通１次学力試験外国語の聴解試験についての調査研究報告書 』，大学入試センター，1985.

9）石塚智一・小野博・清水留三郎・諏訪部真・白畑知彦. 英語リスニングテストに関する実証的研究，大学入試センター研究紀要，No23，１-36，1994.

10）本稿で用いている「妥当性」とは，テスト理論（Test Theory）で用いられている学術用語であり，「本来求めている測定対象となる構成概念を，その測定を意図して行われたテスト得点が，実際に表現していることを支持する度合い」（日本テスト学会『テスト・スタンダード』）と定義される.

11）「内田照久・中畝菜穂子・石塚智一.『大学入試センター試験の英語問題で測られる学力とリスニング・テストで測定される言語運用能力の能力推定値を介した関係性の検討-試験問題の事後標準化とリスニング・テストでの問題提示順序の影響の検討と共に-』，大学入試センター紀要，No33，29-63，2004」，「内田照久・中畝菜穂子・荘島宏二郎.『英語リスニング・テスト実施時に各種騒音が与える影響』，日本テスト学会誌，No１，117-127，2005」，「内田照久・大津起夫・石塚智一.『英語リスニング・試行テストの実施経過と受聴機器選定のためのアンケート調査結果』，大学入試センター紀要，No35，１-18，2006」，「内田照久・大津起夫・椎名久美子・林篤裕・伊藤圭・荘島宏二郎・杉澤武俊.『個別音源方式による英語リスニングテストの予行実施調査』，日本テスト学会誌，No２，41-47，2006」，「内田照久・杉澤武俊・椎名久美子・大津起夫・荘島宏二郎・林篤裕・伊藤圭.『リスニング・モニター試験と改良版ICプレーヤー試作機の実地検証調査』，大学入試センター紀要，No36，１-29，2007」などが挙げられる.

12）大学入試センター.『平成20年度大学入試センター 試験受験案内⑳』，2007を参照.

13）初年度の平成18年度試験では「再テスト」と呼ばれていたが，平成19年度より「再開テスト」と呼ばれるようになった.

14）朝日新聞（2006年１月26日付朝刊），朝日新聞（2007年１月21日付朝刊）を参照．また，平成19年度試験で，不具合の申し出のあった機器を検証した大学入試センターの報告によると，不具合の主な原因は，「再生ボタンの長押し失敗」と「受験生が不具合と受け止めたと考えられるもの」であり，全体の９割を占めているとされる（大学入試フォーラム，No30，66-67，2007.）

15）東北大学高等教育開発推進センター，『テストって何だろう』東北大学高等教育開発推進センターアウトリーチプログラム（1），2008.

16）シンポジウム参加者全体の正確な人数は把握していないため，配布した資料等より概算した参加者人数である.

17）詳細は，「倉元直樹．センター試験リスニングテスト体験-テストの理論と技術 - ，『テストって何だろう』東北大学高等教育開発推進センターアウトリーチプログラム（1），３-10，2008.」を参照.

18）西郡大・木村拓也・倉元直樹．東北大学のAO入試はどう見られているのか? -2000～2006年度学部新入学者アンケート調査を基に-，東北大学高等教育開発推進センター紀要，No２，23-36，2007.

19）Mandler，G.，&Sarason，S. B. A study of anxiety and learning. *Journal of Abnormal and Social Psychology*，47，166-173，1952.

20）藤井義久．テスト不安研究の動向と課題，教育心理学研究，No43-４，455-463，1995.

21）例えば，「Sarason，I. G. Test anxiety and intellectual performance. *Journal of Abnormal and Social Psychology*，66，73-75，1963.」，「松田伯彦･松田文子. 児童の記銘学習におよぼす動機づけとテスト不安の効果，教育心理学研究，No16-２，111-115，1968.」など.

22）Spielberger，C. D. Theory and research on anxiety. In C. D. Spielberger（Ed.）*Anxiety and Behavior*. ３-20. New York: Academic Press，1966.

23）藤井義久，前掲書.

24）大学入試の文脈におけるこうした動機について「西郡大・倉元直樹．日本の大学入試をめぐる社会心理学的公正研究の試み -『AO入試』に関する分析-，日本テスト学会誌，No３，147-160，2007」で，「利

己心モデル」に位置づけられるものとした.

25) Shuler, h. Is there a dilemma between validity and acceptance in the employment interview? In B. N. Nevo&Jager (Eds.), *Educational and psychological testing: the tasktaker's outlook*, 239-250. Tront, Canada: Hogrefe and Hunber.

資料

アンケートにご協力ください

1．リスニング試験に関するアンケートです.

2．最初の講演者（倉元）の話の後に回答してください.

3．結果は全て統計的に処理されますので，学校や皆さんにご迷惑をお掛けすることは一切ありません．感じた通り，率直にお答えください.

1. 次のうち，該当するもの１つに〇をつけてください

性別　（　男性　・　女性　）

属性　（　高校生 [　　年生]　・　教師　・　保護者　・　その他　）

センター試験の英語を受験しようと考えていますか？　（　はい　・　いいえ　）

2. 次の経験の有無について「ある・ない」のどちらか１つに〇をつけてください

2-1. 海外（英語圏）に住んでいたことがある（　ある　・　ない　）

2.2. 学校以外の英会話教室や海外留学などで英会話を勉強したことがある（　ある　・　ない　）

2-3. IC プレーヤーでのリスニング試験（模擬試験を含む）を経験したことがある（　ある　・　ない　）

3. 高校で学ぶ「英語」について，どのように感じているかをお答えください

	そう思わない	あまりそう思わない	どちらとも言えない	少しそう思う	そう思う
3-1. 「英語」は得意だ	1	2	3	4	5
3-2. 「英語」は好きだ	1	2	3	4	5
3-3. 英語教育は，「聞く」「話す」などのコミュニケーションを重視したものがよい	1	2	3	4	5
3-4. 入試での英語もコミュニケーションを重視した試験がよい	1	2	3	4	5

3-5. あなたが思う英語のコミュニケーションとは何ですか（自由記述）

_____

4. センター試験に限らず，リスニング試験一般について，どのように感じるかをお答えください

	そう思わない	あまりそう思わない	どちらとも言えない	少しそう思う	そう思う
4-1. リスニング試験は，「聞く力」を測定するために妥当な試験だと思う	1	2	3	4	5
4-2. 「聞く力」を身につける上で，リスニング試験の勉強は役に立つと思う	1	2	3	4	5
4-3. 筆記によるアクセントや発音の問題でリスニング試験の代用ができると思う	1	2	3	4	5
4-4. 日ごろの努力次第で，リスニング試験の得点は上がると思う	1	2	3	4	5

裏面もあります

	そう思わない	あまりそう思わない	どちらとも言えない	少しそう思う	そう思う
4-5. 国内でどれだけ努力しても帰国子女や海外留学の経験者には，リスニング試験では勝てないと思う	1	2	3	4	5
4-6. リスニング試験がなければ，リスニングの勉強には興味が無い	1	2	3	4	5
4-7. リスニング試験は，自分にとって有利だと思う	1	2	3	4	5

5. **自分が受験するリスニング試験に次のような問題が生じたとき，あなたはどのように感じますか**

	そう思わない	あまりそう思わない	どちらとも言えない	少しそう思う	そう思う
5-1. 周囲の机の音や IC プレーヤー独自の少々のノイズは仕方がないと思う	1	2	3	4	5
5-2. 故障した IC プレーヤーに当たるのは，運が悪かったと納得できる	1	2	3	4	5
5-3. 故障した IC プレーヤーに当たり，再開テストを行ったとしても不利に感じる	1	2	3	4	5

6. **本日の IC プレーヤーでの試聴を通して，どのように思うかをお答えください**

	そう思わない	あまりそう思わない	どちらとも言えない	少しそう思う	そう思う
6-1. 一度も操作の練習をせずに，本試験を受けるのは不利だ	1	2	3	4	5
6-2. 操作においては，何も問題なく上手に使いこなせる自信がある	1	2	3	4	5
6-3. 操作に気をとられて，試験に集中できないかもしれない	1	2	3	4	5
6-4. 周囲の雑音が気になって，試験に集中できないかもしれない	1	2	3	4	5
6-5. 周囲の受験者もほとんど自分と同じように聞こえていると思う	1	2	3	4	5

7. **次の 2 つのうち，どちらが公平・公正なテストだと思いますか．1 つだけ数字に〇をつけてください**

Ⅰ．試験の内容的な妥当性は高いが，受験生全員の試験実施条件を完全に均一にするのが難しいテスト

Ⅱ．試験の実施条件は完全に均一性を確保できるが，試験の内容的な妥当性が低いと思われるテスト

8. **受験者にとって重要であるセンター試験のような大人数が受験するリスニング試験の実施様式の中で，次のうちから，最も適当なものだと思うもの 1 つに〇をつけてください**

Ⅰ．個別音源方式（IC プレーヤー）による試験

Ⅱ．教室のスピーカーによる一斉放送試験

Ⅲ．筆記によるアクセント・発音の問題による代用

Ⅳ．リスニングに関する試験は必要ない

9. **リスニング試験について，何か思うところがあればご自由にお書きください**