

# 第1回 医学教育研究技法ワークショップ報告書

日時 :平成 16 年 11 月 26 日 (土)・27 日 (日)

会場 :東京大学医学部

主催 :日本医学教育学会 医学教育研究開発委員会



日本医学教育学会 医学教育研究開発委員会

## はじめに

平成 15 年夏、医学教育開発委員会を立ち上げる提案をしました。しかし、実際に医学教育研究の質、量を改善していくための方法は決して簡単ではなく、あっという間に時間が過ぎてしまいました。また、その目標を達成するためには、研究の目的、方法論の優劣といった本質的な点についても様々な議論が可能であり、それは、教育心理学、教育学といった関連領域においても比較的似たような問題が起こっているようだということも分かりました。

平成 17 年 11 月末、ようやく第 1 回医学教育研究技法ワークショップを開催する運びとなり、どういう方法論が最善かという点については諸説あるものの、一定の方針、考えを示しながら進めようという考えがまとまっていきました。また、ワークショップに参加していただく先生方に意見をいただき、今後の改善への糸口を探ろうと考えました。本報告書はそういったプロセスをも含めた集大成と言えるものであり、とりあえずこういった考え方が一つあるということを知りやすくしているものであると思っております。

このような試みが実践にはあまり役立たず、学問のための学問になってしまっていないかという不安はありますが、今後とも関係の皆様方から色んな知恵を拝借し、改善に結び付けていきたい所存です。当報告が皆様の医学教育研究の一助となれば幸いです。

平成 18 年 1 月吉日

東京大学医学教育国際協力研究センター 大西弘高

## 目次

はじめに .....	1
参加者一覧 .....	3
プログラム .....	4
開会挨拶 .....	5
講演:医学教育研究の概要説明(大西弘高) .....	7
グループ討論1—研究デザイン .....	19
グループ討論2—データと解析の種類 .....	24
講演:医学教育研究の現状と課題(大西弘高, 足立拓也) .....	27
特別講演:教育研究を推進するために(市川伸一) .....	37
グループ討論3—統計学的解析 .....	58
講演:統計学的解析について(大西弘高) .....	64
ワークショップ後評価 .....	68
参考図書 .....	71

## ワークショップ参加者

グループ	名前	所属
1	有田 彰	埼玉医科大学短期大学 臨床検査学科
1	大滝 純司	東京大学医学教育国際協力研究センター
1	樫田 美雄	徳島大学総合科学部
1	岸 太一	東邦大学医学部 心理学・教育開発室
1	黒岩かをる	薫陶塾
1	西城 卓也	名古屋大学病院総合診療部
1	水野 洋子	東京大学病院神経内科
1	安田 幸雄	金沢医科大学医学教育学
1	渡邊 淳	日本医科大学付属病院遺伝診療科(生化学第2)
2	伊木 雅之	近畿大学医学部公衆衛生学
2	井上千鹿子	立教大学
2	大生 定義	横浜市立市民病院神経内科
2	片岡 竜太	昭和大学歯科病院 顎口腔疾患制御外科学
2	佐藤 淑子	福岡大学医学部教育計画部
2	鈴木 康之	岐阜大学医学部医学教育開発研究センター
2	谷口 純一	熊本大学医学部附属病院 総合診療部
2	中島 昭	藤田保健衛生大学医学部 医学教育企画室
2	三原 祥子	東京女子医科大学医学部日本語学
2	湯浅 孝男	秋田大学 医学部保健学科 作業療法学専攻
3	石井 誠一	東北大学大学院 医学系研究科 医学教育評価センター
3	岩堀 禎廣	薫陶塾
3	小田 康友	佐賀大学医学部附属病院総合診療部
3	菅野 義彦	埼玉医科大学 医学教育センター・腎臓内科
3	木村 琢磨	国立病院機構東京医療センター 総合診療科
3	丹羽 雅之	岐阜大学医学部医学教育開発研究センター テュートリアル部門
3	堀 有行	金沢医科大学医学教育センター
3	守屋 利佳	北里大学医学部 医学教育研究部門
3	吉井 文均	東海大学 神経内科
3	吉田 一郎	久留米大学医学部医学教育学

## 医学教育研究開発委員会

委員長	大西 弘高	東京大学医学教育国際協力研究センター
委員	相野田紀子	金沢医科大学医学教育学
〃	犬塚 裕樹	久留米大学医学部医学教育学
〃	杉本なおみ	慶応義塾大学看護医療学部
〃	田川まさみ	千葉大学卒後臨床研修部
〃	松尾 理	近畿大学医学教育センター
〃	森 敏昭	広島大学大学院 教育学研究科
〃	吉田 素文	九州大学医療系統合教育研究センター
講演	足立 拓也	ダンディー大学医学教育センター大学院生
記録	長田 真美	東京大学医学教育国際協力研究センター大学院生

## 第1回医学教育研究技法ワークショップ

日時:平成17年11月26日(土)1400-11月27日(日)1230

場所:東京大学医学部国際協力研究センター医学部総合中央館(医学部図書館)310室

一般目標: 医学教育研究の実行が可能な実践的手法を習得する

- 個別目標
- ・ リサーチクエスチョンに対し、自ら研究計画を立てられる
  - ・ 研究立案、実施の流れについて生物医学的研究との違いを列挙できる
  - ・ データに対し、適切な統計解析方法を述べることができる

### ・第1日目

- 14:00-14:30 開会挨拶、自己紹介
- 14:30-15:00 概要説明:医学教育研究のデザイン、サンプリング(14:30-15:15に変更)
- 15:00-15:45 グループ討論1:シナリオに沿った研究経過—デザイン、サンプリング  
(15:15-16:00に変更)
- 15:45-16:15 全体発表と討論 6分間×3グループ(16:00-16:45に変更)
- 16:15-16:30 休憩(16:45-16:50に変更)
- 16:30-17:15 グループ討論2:シナリオに沿った研究計画—データと解析の種類  
(16:50-17:35に変更)
- 17:15-18:00 全体発表と討論(17:35-18:10に変更)
- 17:35-18:15 講演:医学教育研究の現状と課題(18:10-18:50に変更)  
東京大学医学教育国際協力センター 大西弘高先生  
ダンディー大学医学教育センター 足立拓也先生
- 18:15-18:45 全体討論(18:50-19:00に変更)
- 18:45- フォレスト本郷へ移動(19:00~に変更)
- 19:45- 懇親会(自由参加、会費別)

### ・第2日目

- 9:00- 9:45 講演:教育研究を推進するために —教育心理学的方法をどう生かすか  
東京大学大学院教育学研究科教育心理学コース教授 市川伸一先生
- 9:45-10:30 質問と全体討論(9:45-10:40に変更)
- 10:30-10:45 休憩(10:40-11:00に変更)
- 10:45-11:30 グループ討論3:仮想データに対する統計解析方法(11:00-11:45に変更)
- 11:30-12:15 全体討論と発表(11:45-12:30に変更)
- 12:15-12:30 閉会(12:30)

## 開会挨拶(大西先生)

皆さん、今日はお集まりいただきましてありがとうございます。医学教育学の分野は非常に急成長しています。これは日本でも医学教育改革が進んでいるということもありますが、世界的にもやはりこの10年くらいめざましいものがあります。その世界的な原動力になっているのが疫学教育研究ではないかと思われまます。そのような思いもあって2003年医学教育学会では医学教育研究を専門に扱う専門委員会「医学教育開発委員会」を立ち上げました。8名の委員でやってきましたが、ルールがないところを試行錯誤でやってきました。この年末で任期が終わりますので、一度ワークショップをするということで、我々の方も持っている知識や技法とかを整理し、いろんなもの、経験を共有しながら学んでいきたいと思ひます。

全体討論およびグループディスカッションの場所、設備の説明。

## 自己紹介(参加者全員)1名30秒間

### グループ3

小田先生:佐賀大学医学部総合診断部、臨床教育と医学部全体のカリキュラムの改革を担当している、今日は勉強したいと思ってきました

守屋先生:北里大学医学部 医学教育研究部門、腎臓が専門、教育のアウトカムをどのように出していくか

菅野先生:埼玉医科大学・腎臓内科、今年の5月から医学教育センター

吉田先生:久留米大学の医学教育、アメリカで質的研究を勉強しましたが、よくわかりませんでしたので、きました。

丹羽先生:岐阜大学、研究となるとよく分からない状態が続いています。学んで活かしていきたい

吉井先生:東海大学神経内科、教育計画部、現在SP教育に力を注いでいます。データはたまっているのですが、いかに科学的に分析していくのかが分かっていません

堀先生:金沢医科大学医学教育センター、もともと神経内科、そろそろ医学教育で考えていかなければいけないと思っています

石井先生:東北大学、昨年から教育評価センター、カリキュラムが変わるのですが、なぜ変えるかを客観的にいえるものがなく、経験則でやっている。

木村先生:国立病院機構東京医療センター 総合診療科、研修医指導、地域医療保健研修の仕事をしています、なかなか研究にいたらないので勉強したい

岩堀先生:薫陶塾、医療者のコミュニケーション、SPを使って研修、結果や効果が出るその根拠、より効率の良いものへ、データ処理を探りたい。

### グループ2

湯浅先生:秋田大学医学部保健学科、作業療法士、教育もまだまだ、養成校作業療法教育研究会の企画担当、レベルアップを、他職種に負ける危機感をもって参加

鈴木先生:岐阜大学医学部、研究センターとして5年目、スタッフは自然科学をベースにしている、

教育研究が課題、

中島先生：藤田保健衛生大学、生理学、現在主な仕事は学生の入試をどうするか、入学後の学生をどうするか、

大生先生：横浜市立市民病院神経内科、この病院で16－20名の学生の研修責任者を担当、

谷口先生：熊本大学 総合診療、研修医、指導医養成に関与、6年前に大学に戻った、医学教育、研究的なことを勉強したい

三原先生：東京女子医大、日本語、文系、医学部と看護学部の1年生の表現法に関わる

井上さん：立教大学学生、日大でSPをやっていた。行動科学の情意教育に興味がある。

伊木先生：近畿大学、公衆衛生学、医学教育研究センター、医学教育への疫学的研究ができるか

片岡先生：昭和大学歯科病院、PBL 委員会所属、歯科ではPBL 導入して日が浅い。反対派多い

佐藤先生：福岡大学医学部教育計画部、データが溜まってきました。そろそろ処理を

## グループ1

安田先生：金沢医科大学医学教育学、5年 学生教育ばかりをやっている、実績をあげるべく出席

岸 先生：東邦大学医学部心理学・教育開発室、4 月から教育開発、医学教育を専任で担当するのは今年から

黒岩先生：薫陶塾、発足7年、データは貯まるばかり、SP 教育、教育効果の高い研修を目指して

西城先生：名古屋大学総合診療部、・・・

渡邊先生：日本医科大学附属病院遺伝診療科、生化学、臨床遺伝教育を医学教育に展開させるか

有田先生：埼玉医科大学短期大学、臨床検査学科、医学部以外ぜひ皆さんのお知恵を

樫田先生：徳島大学総合科学部社会学を専門、医学教育を専門、医学のメソドロジー、VTR 分析、オスキーやPBL を研究

大滝先生：東京大学医学教育国際協力研究センターから 2 ヶ月前に移動、今東京医大総合診療の立ち上げをしています

委員の先生の自己紹介

概要説明(大西先生)

いろんなアプローチがあるだろうと思います。各自テーマを出し合って、そのテーマを取り上げてディスカッションしてもいいのですが、研究はその展開と競合・共有のバランスが難しいと思われれますので、ある枠組を提示して医学研究の進めていけるという枠組を、それに対する皆さんの考えをきいて展開していく形にしました。そこで、この枠組で提示するそのなかで展開し、この中には心理学、社会学もいらっしやる。かなり科学的な眼で研究を進めている方もいらっしやると思いますが、基礎医学を科学的な考え方で見ている方もいらっしやあって、研究をすすめている方も多いと思います。医学研究と教育研究との違いなどについて概要を説明します。

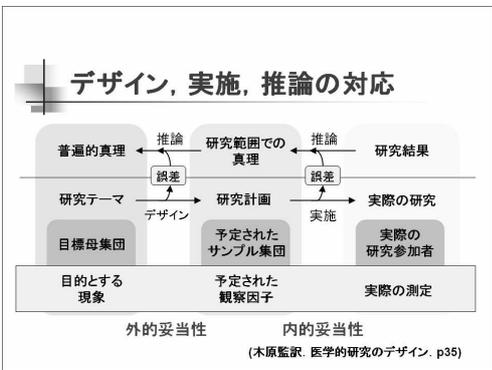
**1、教育研究の流れ:**これだけで研究として成り立つとはいえないけれど、この部分がしっかりしていなければいけません。研究デザイン→内的妥当性をクローズアップすることによって違いが見えてくるのではないかという印象をもちました。サンプリング→代表性や一般化可能性を考えるうえでサンプリングの問題は欠かせないだろうと思われれます。心理学でよく使われる構成概念の考え方ですね。測定したいもの

**教育研究の流れ**

1. 研究デザイン: 内的妥当性への脅威
2. サンプリング: 代表性や一般化可能性
3. 測定する構成概念
4. 変数, 信頼性と妥当性
5. データ回収
6. 統計学的解析(→2日目の話題へ)

というものが、器械や測定器具みたいなもので当てると測定できるというものはないわけです。ある場合には観察を通して、ある場合にはそれぞれの学習者のふり返りを通して数値化されていく、そういうところがありますので構成概念という考え方を整理してみたいと思います。それから変数ですね。その概念を測定して数値化していったときに変数をどのようなものを選ぶのか。そのとき信頼性妥当性がどのように変わっていくのか、このあたりも重要です。データ回収に関しては、少しだけ注意点を挙げたいと思います。統計学的な解析は明日の話題にしたいと思います。

**2、デザイン、実施、推論の対応:**これは研究の全体的な流れを示したもので、これは医学的研究のデザインという本のなかにかかれていたもので、これが非常に使いやすかったです。まず一番最初に目標母集団、研究テーマというのがあって、例えば非常に大きなテーマであると、日本の医学生のコミュニケーションの能力を高めたいというテーマがあったとしますと、それに対して日本の全体の医学生になり



ますが、しかしデザインを組んで研究をしていくときにはおそらく何らかの予定されたサンプル集団からデータをとって研究計画を立てていく。このサンプル集団というのは大学の中でデータをとる

のが簡単だということで〇〇大学の医学生という形で、例えば東京大学の医学生100人分を測定した時にそれは全国の医学生の代表になるのかそのようなことが問題になるんですね。実際に研究をして、この予定された集団の中から実際に研究に参加してくれる人がどれくらいいるのか、もちろん教育のセッティングでこれは授業だから必ずやらないといけないということであれば、全員が履修して、全員からデータをとることができるかもしれません。がその時に脱落するというのも起こりえます。これらを通じてみていきますと、目的とする現象があってそれから予定された観察因子があり、実際に測定をして、結局実際に測定した結果が出ます。それから研究範囲内の普遍的真理、こういうところに繋げていくわけですけど、ここに結果から研究範囲内のサンプル集団での心理とは何かというのを推論するわけですね。そこには何らかの誤差というのが入ります。あるいはこの研究範囲内の心理、例えば東京大学の医学生でコミュニケーションの能力はどのようなのか、とういことがある程度分かりました、ということ在全国にどのくらい適用できるかという問題がってきます。結局、ここのサンプル集団から実際に研究に参加する、ここには内的妥当性という問題があって、もう一つ、研究範囲から普遍的な真理に移ることができるかどうか、ここには外的妥当性あるいは一般化可能性があります。そういうふうな考え方が必要です。これが一番の研究の骨格になります。ただこれは科学的な研究であってもほぼ同じだと思います。

**3、研究デザイン:** 研究デザインというものにはどういふものがあるのか。まず事後テスト単独と事前事後テストがあります。これを比較しますと事前事後テストによって改善度が測定可能になります。それからもう少し複雑なデザインになりますと、対照群を設ける。対照群がありますと、準実験と呼ばれますけど、もう少しいろいろな要因が統制できるようになります。さらにランダム化の割り付けをするかですね。ランダム化していますと真の実験とよべれます。これによってさらにいろいろな要因が統制されます。

### 研究デザイン

- 事後テスト単独 vs 事前事後テスト
  - 事前事後テストにより改善度が測定可能
- 対照群の有無(対照群有: 準実験)
- ランダム化の有無(ランダム化: 真の実験)

**4、一般的な研究デザイン:** 一般的な研究デザインをみてみますと実験的なものではない前実験的、それから準実験的、真の実験、この3つのレベルに分かれています。事後テスト単独ですとより事前事後テストは前後の比較ができ、改善度がわかります。真の実験の中にもランダム化してグループを設けて事後テストだけのものと事後テスト両方のもの、こちらの方がコストが安く済みますし、ランダム化がうまくいってればこれだけでも十分通用するわけですけど、ランダム化がうまくいっているのかを確認するだけでもなかなかこれだけでは難しい、という問題もあります。これがかなり完璧なデザインといえるのではないかと思います。ランダム化していないこちらのほうはどうかと申しますと結局このグループとこのグループがいろいろな要因によって、等しいグループかどうか分からないですね。たとえばコミュニ

### 一般的な研究デザイン

前実験的	事後テスト単独	X -- O
	事前事後テスト	O <sub>1</sub> -- X -- O <sub>2</sub>
準実験的	非ランダム化比較 事前事後テスト	E O <sub>1</sub> -- X -- O <sub>2</sub> C O <sub>1</sub> ----- O <sub>2</sub>
	ランダム化比較 事後テスト単独	R E X --- O <sub>1</sub> C --- O <sub>1</sub>
真の実験的	ランダム化比較 事前事後テスト	R E O <sub>1</sub> -- X -- O <sub>2</sub> C O <sub>1</sub> ----- O <sub>2</sub>

X: 教育介入, O: 観察か測定, E: 実験群, C: 対照群, R: ランダム割り付け

ケーションの能力を比べたいという場合に、実験群のほうには非常に女性が多かった、こちらのほうには男性が多かったといった場合に、こちらの実験群の方が点数が高いというのが性別による要因なのか、あるいは何らかの教育効果なのかかわからないということが起こってきます。

**5、前実験的デザインの弱点:**このあたりのことを整理しますと、前実験的の弱点として、例ですが、生理学の授業に関する効果の確認を試みました。事前事後テストデザインでした。事前は55%、事後テストは1カ月後70%でした。授業は有効だったと結論できるか、ということなのですが、この時に内的妥当性のいろいろな要因を考えたほうがいだろう。Threatという英語であらわしていますが、日本語で脅威といいます。英語

**前実験的デザインの弱点**

- 例:生理学の授業に関する効果の確認を試みた
  - O<sub>1</sub>--X--O<sub>2</sub>:事前事後テストデザイン
  - O<sub>1</sub>:55%, O<sub>2</sub>:70% (1ヶ月後)
  - 授業は有効だったと結論できるか
- 考慮すべき脅威 (Threat)
  - ✓ 状況効果(学習者に外因が与えた影響)
  - ✓ 成熟効果(時間経過による変化)
  - ✓ 測定効果(測定者や測定法の変化)
  - ✓ テスト効果(1回目のテストが2回目を改善)

ではヒストリーというのですが、状況効果、学習者に外因が与えた影響によってこの差が出てくる。例えばこのときに日本人でノーベル生理学賞をとった人がいて、生理学に関心の高い人が多かった、その影響を受けた可能性がありますと、この授業の効果だけではなくて全然関係のないところで起こったことによって点数が影響を受けたこととなります。あるいは時間経過によって影響が出てくる。当然元々ある理解が深まる。生物を取った人が元々知識があるということもありでしょう。それから測定効果で測定自体をきちっとできていたのか、測定者が変わるあるいは測定法が安定しないということで勝手にこういう差がでてきただけという場合もあります。テスト効果といって一回目テストをやっていますからその時にある程度テストに対する対応の方法も覚えますし、学習もしますから、そういうことで1回目のテストより2回目は改善するということはありません。このようなことをいろいろ言い出しますとこのぐらいのデザインではほとんど何もいえないでしょう、という感じが段々見えてくると思います。

**6、準実験デザインの弱点:**続いて準実験デザインです。非ランダム化の比較で事前事後テストデザインを用い、地域の患者を継続的に訪問するプログラムを1年次に開始しました。今年の2年次はこのプログラムはやっていない。これに対して1年次に対して患者の問題に全人的に対応できるかその能力を口頭試問(がうまくいくかどうかはさておいて)で評価した。1年次は5段階評価で4.2をとりました。2年次は3.1で有意差がみられました。しかし、今回の1年次は女子の率が上がりました、それに平均年齢も高かった。おまけに、今年面接方法を見直して一人ひとりに時間をかけるようにした。このようなことがありますと、そういった要因、性別による要因、年齢による要因あるいは面接方法によっていい学生が入ってくるようになったかもしれないから、そういった要因がこの点数差に表れているだけかもしれません。それからもう一つホーソン効果も準実験デザインの場合、大きな問題。このように比較をしている時に教育介入されている群は明らかに自分たちは注目されていると分かっていますから、それに対して意識が高まってい

**準実験デザインの弱点**

- 今年から地域の患者を継続的に訪問するプログラムを1年次に開始。今年の2年次、1年次に対して患者の問題に全人的に対応できる能力を口頭試問で評価
- E O<sub>1</sub>--X--O<sub>2</sub> 非ランダム化比較
- C O<sub>1</sub>-----O<sub>2</sub> 事前事後テストデザイン
- 1年次は5段階評価で4.2、2年次は3.1で有意差みられた
- しかし、今回の1年次は女子の率が高く、平均年齢が高かった。おまけに、今年から入試時の面接方法を見直したところだった
- ホーソン効果

るわけですね。ですからテストに対してもより準備しようという心理状態になるといわれています。

**7、各デザインの主な利点・欠点:** そういうことで、今いった5つのデザインに対して、利点欠点というのをあげました。結局、前実験的なデザインでは状況効果、成熟効果、テスト効果、測定効果など考慮を要します。しかも事前事後でない介入の変化も不明である。こちら(前実験的)の方は介入前後の変化は分かるけれど他の効果は同様です。それから準実験的なものに関しては、状況効果、成熟効果、こういったことは排除は

前実験的	X--O	状況効果、成熟効果、テスト効果、測定効果は要考慮。介入前後の変化不明
	O <sub>1</sub> --X--O <sub>2</sub>	介入前後の変化は分かるが他の効果は同様
準実験的	E O <sub>1</sub> --X--O <sub>2</sub>	状況効果、成熟効果、テスト効果、測定効果は排除可。選択バイアスは要考慮
	C O <sub>1</sub> -----O <sub>2</sub>	
真の実験的	R E X ---O <sub>1</sub>	選択バイアスは排除可。介入前後の変化不明
	C ---O <sub>1</sub>	ランダム化が上手くいったか不明
	R E O <sub>1</sub> --X--O <sub>2</sub>	ランダム化が上手くいったかどうか事前評価可能。対照群は教育介入受けない
	C O <sub>1</sub> -----O <sub>2</sub>	

可能である。でも選択バイアスは要考慮。このグループを分けるときにきちっと分けられているかということで、教育介入を受けたい人、受けたくない人ということで分けますと明らかにいろんな差が出てきます。先ほどのように1年次、2年次ということで行きますと、そういった自分の好みによっての影響はないわけですが、ほかのいろんな要因がからんでいきます。それから、真の実験的デザインですね、この事前事後になっていない事後だけのものに関しては、選択バイアスは排除が可能である。でも、介入前後の変化は不明ですし、ランダム化がうまくいったかどうかはわからなくて、ここのところで結果が出てきても、ランダム化がうまくいっていないだけかもしれません。こちら(ランダム化比較事前事後デザイン)の方はそういった意味ではかなり完璧に近いですね。但し、こちらのコントロールグループの方は教育介入 X を受けていないという問題が出てきます。こういうことを考えてみますといくつか難しい点が出てきますが。

**8、更に難しい点:** 対照群を設ける時に、対照群は教育しないということになりますからこれはいいのか、あるいはその教育しないよりはした方がいいというふうな気持ちは誰にでもあると思いますが、しない群とする群とランダム割り付けをするというのはいいのかよくないのか。教育現場の倫理的な問題ですがなかなかその辺が難しい。それから、カリキュラムが研究の対象になることが多いのですが、カリキュラムを実験的に

<ul style="list-style-type: none"> <li>■ 対照群を設けること(教育しない群ができる)、ランダム割り付けすることは倫理的にOK?</li> <li>■ カリキュラムは試験的に変更するのか?</li> <li>■ 対照群は人口動態学的要因、心理学的要因に関して実験群と同等か</li> </ul>
→ 時に準実験デザイン、真の実験デザインを選択すること自体が困難

研究のために変更するのか、その変更というのは一時的なものなのか、あるいはその何らかのカリキュラム導入プロセスの一貫なのかそういったことも、正式に導入するとすると学生もそれなりに準備をしていきますから、いろんな影響が出てくる。それからランダム化していない場合、対象群というのは人口動態学的要因、あるいは心理学的な要因に関して実験群と同等かどうか、ここは比較試験だと常に問われるところですが、いろんな要因に関して調べておいてきちっとランダム化ができていだろうか確認する必要がある、いろんな研究で行われています。そのあたりはおそらく疫学の場合でもほとんど同じことになるでしょう。こういったことをいろいろ考えて、準実験デザイン、真の実験デザインをとりたいけれども教育現場だとそういうことがなかなかうまくいかないあるいは許されない、といったことも起こりえます。今のデザインのところで何か質問はないですか？

Q: 更に難しい点の1番目のところですが、一番上の項目なんです、教育する群としない群がどうしてもでてしまう。たとえば新しくやろうとしているプログラムが効果があるかないか分からないので、問題ないという言い方で倫理委員会を通ると聞いたことがあるんですけども、それは本当でしょうか？国際的にそういうのは認められるのかなと思ひまして。

A: 去年のプログラムと今年のプログラムを比べるということですね。さきほどの例でいうと X が入っているのと入っていないということですね。X1 と X2 という形のデザインになるということになりますよね。その方が倫理的には安全性が高いと思われま。認められる可能性は高いと思われま。それだと何と何を比べているのかということが非常に曖昧になりますよね。要するに去年の教育に比べて、今年教育がいいのかどうかということが、去年の教育と今年教育をうまく定義してそういうことを論文にして、それが一般的に受け入れられるのかどうか、他の大学にとってはそういう古いカリキュラムでやっていないという場合はどうか。例えば講義中心を PBL に変えるとか、これはかなりユニバーサルな動きなので、そういう変化だと比べても結構意味があると思ひま。すけど、例えばコミュニケーショントレーニングの新しい試みをやりましたという場合には古い方法と新しい方法の比較になり、古い方法自身があまりユニバーサルでないと他に役に立たないという事はあと思ひま。そのあたり難しいですけども。実はこの O2 という測定の後にですね、こちらのコントロールグループに教育プログラムを設けるという抜け道はあります。ここに X を設ける、だから何かのプログラムの途中で半分くらいの学生にこのプログラムをやったときに一斉にテストをやる、という方法はなくもないです。但し学生からは教育介入を受けていないのにテストをされるなんてといった拒否反応は必ずあります。

Q: 一番下のランダム化をやればそれは勿論良くなると思ひますが、実際にそのランダム化というのはどの位の項目までこだわってできるものなんでしょうか、例えば医学部の学生、普通の学生にしてもクラブ活動だとか住所だとかその辺までも、もしかしたら通学時間までもありますし、そういうところまでランダム化しないともは言えないのか、もっと単純に去年の成績をとって見ましたということでそのくらいで、勿論そのいうテーマにもよると思ひま。すけど現実的に僕はあまり論文読まないのだからわかりませんが、論文でされているのはどの変まででしょうか？

A: まず、マッチングという(はいそうです)、ペアマッチをするということとランダム化するということは違います。だからいろんな要因を分けてそのときにマッチングするということはあるん。沢山の要因はできない、せいぜい2つ3つくらいですね。ランダム化というのはとにかく乱数表などを使って割り付ける、その後でそういったクラブ活動とかのこともですね、全部比較ができると思ひま。す。但し、そうやってランダム化やってみたけれども50人:50人で実はこちらのグループはずいぶん男性が多かったとそういったことになる場合もあります。そういう場合には、研究を続けていいのかどうか難しい課題にはなります。

Q: 言葉の定義ですけど、妥当性についての呼び方使い方でけっこういろいろ使い方が違っていた

ことがあるんですけど、この今の木原先生の訳したやつの内的妥当性と計画が実際と比べてどうかということなんですけれど、どれくらいそこがっているかということと・内的妥当性と呼ぶのかという説明だったと思いますが[ハイ]、僕のイメージはたとえばあるものを測ろうとしていてそのあるものの構成概念はまた別だとして、これを測ろうとしているものがはっきりしてそれを測る方法が妥当かどうかというのめっこう内的妥当性という言い方でいうことが多いようなイメージをもっていたんですが、ここちょっとニュアンスが違うかなと思って言ってみたんですけどその辺のところはどうなんでしょう？

A:これは研究デザインの内的妥当性で、内的妥当性という言葉自体は研究デザインで使うものだと思います。ただ妥当性という言葉自体は、測定の構成概念がきちっと測定できているかどうかそのこと自体を妥当性といいますので、それとこれとはまたちょっとレベルが違う。

Q:測定の妥当性については普通よばないということ？

A 測定の妥当性についてはまた後でできます。

それではサンプリングにいきます。

**9、サンプリング:**サンプリングはランダムにやるとかあるいはたとえば確率的じゃない方法でやるといったこの2種類があります。なかなか教育現場でこのクラス全体ということが多いようですが、あるいは全国調査をやる、全国に郵便でばら撒いてやるという形が多いのかなと思います。大体そういう場合には、連続してとる連続サンプリングという形になっているのかなと思います。確率的な方法の中にはいろんな系統的な、層化する、例えば地域でわけるとか、あるいは人口密度で分けるとかいろんな方法があると思います。ただ連続サンプリングというのが非確率的なサンプリングのなかでは一番いいといわれていて、クラス全体というのは、そういう意味ではある意味では安全性がある、教育現場では複雑なサンプリングが困難な面があります。

**10、研究の規模と一般化可能性:**ただここは研究の規模と一般化可能性に話がかかってきます。

学内だけでとどめるとか、あるいは国内、あるいは国際比較、1学年のみにするかあるいは複数学年でやるとか…いろんな方法がありますが、どの程度一般化可能性の結論を導き出したかによって考えていくわけですけど、当然研究実現性を考えるとあまり手を広げられない、というのが現状です。

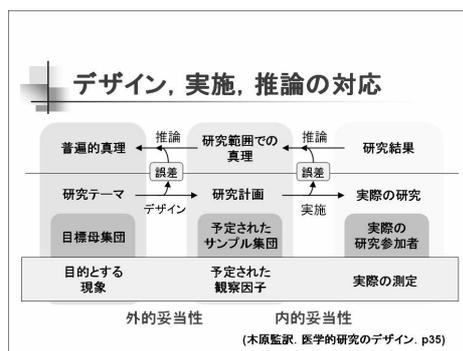
### サンプリング

- 確率的サンプリング
    - 単純ランダムサンプリング
    - 系統的サンプリング:何らかの周期性による
    - 層化ランダムサンプリング:各サブグループから
    - クラスタサンプリング:区域を選んでから
  - 非確率的サンプリング
    - 連続サンプリング:所定の人数まで連続取り込み
    - 簡便サンプリング:サンプリングしやすいところから
    - 選別サンプリング:研究しやすい対象者から
- 教育現場では複雑なサンプリングは困難

### 研究の規模と一般化可能性

- 学内, 全国, 国際比較...
  - 一学年のみ, 複数学年...
- どの程度一般化可能な結論を導き出したか、研究の実現可能性とを併せて考慮することに

**11、デザイン、実施、推論:** サンプルングというのは、予定されたサンプル集団から研究範囲内の真理が分かります。それが普遍的な真理に結び付くというところでのようなサンプルをとっておけばどのくらいの一般化可能性をもってものが言えるかが決まります。



**12、教育、心理領域の測定の難しさ:** 教育研究でおそらく一番難しいのは、測定がなかなか一筋縄ではいかないということです。測定するためには何らかの構成概念というものを一端仮定し、それを検討しながら使っていく、という形になります。定義するなら構成概念とは、定義の難しい概念を測定するために現象に名前を付けたもの、例にあげた愛、愛とはなんぞやといえば、親密性、情熱、傾倒、とか、ルビンの愛着、援助、排他性、親密さの各欲求というようにいろんな言葉や形で定義するわけですね。そうすると、このそれぞれは測定できて両方うまくからみあわせて愛になりますのでそういう説明になるわけです。そういうもので愛は測れないという人も多いかもしれませんが、これが心理学のやり方なのです。

### 教育、心理領域の測定の難しさ

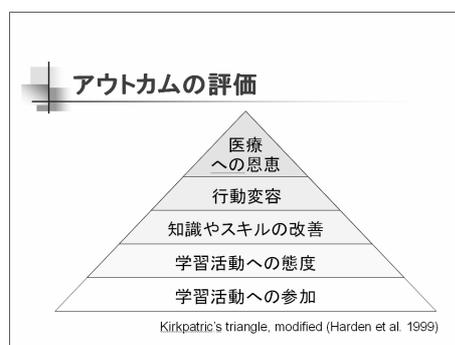
- 測定するためには構成概念が必要
- 構成概念: 定義の難しい概念を測定するために現象に名前を付けたもの
- 例) 愛
  - スタンバーグ: 親密性、情熱、傾倒
  - ルビン: 愛着、援助、排他性、親密さの各欲求

**13、成績とは?:** 成績あるいは能力というのは我々教育現場ではいつも問題になっていますけれど、能力の高そうなひとに高い成績をつけたいというときに、MCQを使って実際に得点が高いかどうかで進級判定をします。こちらの方は非常に客観的な点数であらわれているような気がします。がこんな案でいいのかなという感じもします。一方でやはり専門家が実際に見た方がいいだろうということで推薦状というものすごく曖昧な測定方法ですばらしい内容が書かれてるといいかということ、これもなんだか怪しい気がします。結局こういったことをいろいろ考え合わせてひとつの概念としての評価案をたてていくしかないだろうと思います。

### 成績とは?

- 能力の高い医師になりそうなのは?
  - MCQの高得点
  - 推薦状での素晴らしい内容

**14、アウトカムの評価:** やはり医療でもそうですけど、教育でもアウトカムということがずいぶん問われるようになってきました。一番低いレベルですと、学習活動に参加したかどうかだけで問われるようなレベルがあります、活動への態度はどうか、満足度はどうかあるいは知識やスキルは改善したのか、それからいろんな行動変容したのか、学習者の行動変容だけでなく、それが将来的に医療に恩恵をもたらしたか、このあたりまでいくと



非常に医療に対して貢献した、という感じになってくると思いますが、なかなかこのレベルまでいく研究はありません。

**15、質問紙作成の手順:** 構成概念ということを考えて、これを実際に測定するときが一番よく使われるのが質問紙、ですので、少し質問紙に関して細かい話をします。よくあるのが、質問紙を作る、いわゆるアンケートですね。心理学領域の質問紙というのはちょっと違って、項目をいきなり作るということではいきなり書き始める、これは大体専門家としてやってはいけないことです。まず、その測定目的、あるいは対象構成概念を明確化してい

**質問紙作成の手順**

1. 測定目的の明確化:ブレインストーミング, フォーカスグループ, 深層インタビューetc
2. 項目作成:まずは既存の質問紙探索から
3. パイロット試験:信頼性, 妥当性の検証
4. 項目分析

く、ということでブレインストーミング、フォーカスグループ、深層インタビュー、これらによって質的に情報を集めて、何を測定しようかとかなり明確化させてはじめて項目作成に行ける、ということになります。また項目自体は作成できても、それ自体信頼性があるいいツールになるかどうかわかりません。だから、まず既存の質問紙を探そうというところから始めた方がいいといわれています。何らかの形で新しく作ったとすればそれをパイロット試験して、信頼性妥当性の検証をする、それから項目の分析を行うという手順になります。

**16、質的 VS 量的データ:** 質的なデータ、量的なデータの違いですね、質的研究というのも大きなテーマになってきています。教育現場における質的研究について余り詳しくないのでその話は十分できません。ただこの違いですね、量的データというのは見た目には客観的なデータに感じられるということが一つの特徴です。統計学的な解析が可能なのでそういった意味できちとしたことが言いやすいというところがあります。質的なデータ

**質的 vs 量的データ**

- 量的データ
  - 見た目は客観的なデータにみえる
  - 統計学的解析が可能
- 質的データ
  - 情報量は豊富
  - 結論を出す、仮説を検証することは困難

は情報量が豊富というのが特徴です。但し、これをまとめ上げて結論を出すとか、仮説を検証するということは一般的には困難であるといわれています。一般的にアンケートを作ってもよく数字的に1~5まで選びなさいという部分と、オープンになっているところとを設けておくという形がありますが、それが量的質的にあたると思ってもいいかもしれません。

**17、変数:** 変数ですね。良くある形、X の変化につれて Y がどのように変化するか、いう形の変数をとったときに X が独立変数と呼ばれて Y が従属変数と呼ばれます。例えば、患者満足度によって研修医の診療評価をやりました。患者の学歴が低くて、年齢が高く、研修医が女性の場合には評価が高くなったという場合に、独立変数が患者の学歴であり、患者の年齢であり、研修医の性別である、従属変数は患者満足度得点であるという

**変数**

- Xの変化につれて、Yはどう変化するか
  - X: 独立変数, Y: 従属変数
- 例: 患者満足度による研修医の診療評価
  - 患者の学歴が低く、年齢が高く、研修医が女性の場合に評価が高くなった
  - 独立変数: 患者の学歴, 患者の年齢, 研修医の性別
  - 従属変数: 患者満足度得点

ことになります。これらの要因が得点に影響を与えている、ということですね。

**18、変数の種類:**変数の種類、このあたりは皆さんご存知ですね。名義変数、二区分変数、順序変数、間隔変数があります。二区分変数というのは、二区分のものに特化した分析がありますので、これを分けておいた方が便利です。順序変数のなかには、通信簿の1・2・3・4・5あるいはアンケートのなかで1・2・3・4・5とよく出てきますけど、この1・2・3・4・5の1と2、2と3の間隔というのが等しいかどうか分からない場合、この場合間隔変数ではなく、順序変数になります。一般的には、全くなし、少し、多少、かなり、非常に多い、というものなどが典型的な順序変数ですね。教育現場では間隔変数そのものというのはあんまり出てきません。たとえば授業の出席回数とかは間隔変数ですけど、いわゆるテストの成績が100点満点でこれを間隔変数とよんでいいのかどうか微妙な問題です。

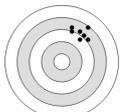
**変数の種類**

- 名義変数
  - 東京都-神奈川県-埼玉県-千葉県
- 二区分変数
  - Yes-No, 合格-不合格
- 順序変数
  - 通信簿の1-2-3-4-5
  - 全くなし-少し-多少-かなり-非常に多い
- 間隔変数
  - 授業の出席回数

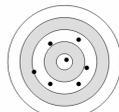
**19、信頼性と妥当性:**一般的には信頼性が測定の安定性、妥当性は測定したいものが測定できているかどうか、ということを目指しています。こちら左図の方は非常に弾痕が集まっていますから信頼性が高い、でも妥当性は低い。真ん中に届いていない。こちら右図の方は信頼性は低いのできちっと測定できているかどうかはかなり疑問なんですけど、全体としては妥当な方向は向いてそうだとことはわかります。先ほどの例では、こちら左図がMCQで能力を測る方、こちら右図が推薦状かもしれません。

**信頼性と妥当性**

- 信頼性: 測定の安定性
- 妥当性: 測定したいものが測定できたか



信頼性は高いが  
妥当性は低い



信頼性は低い  
全体としては妥当な方向

**20、評価の妥当性とエビデンスの分類:**

妥当性について今、Messickという人が1991年に新しい妥当性の定義を出しています。ここにいろんな本をもってきていますけど、この本ですね『教育測定学』のなかにその定義のことが書いてありますが、あまりに哲学的でよくわかりません。非常に難しい、で、教育評価の妥当性というもののエビデンスを示すために、こういう5つの要因に分けた方が分かりやすいのではないかと、これはイリノイ大学のDowningが書いたものです。いろんなテストの内容のこととか、処理プロセス、内的構造という形でエビデンスをそろえていくと妥当かどうか分かりますよ、という

評価の妥当性エビデンスの分類 (Downing, 2004)	
内容	試験のブループリント、達成領域に対するブループリントの代表性、テスト計画書、テスト計画書に対する項目内容の一致、領域に対する項目の代表性、達成領域に対するテスト内容の論理的/経験的な関連性、テスト問題の質の高さ、問題作成者のレベル、感受性検査 <sup>[1]</sup>
処理プロセス	学習者によるテストフォーマットへの慣れ、電子的スキャン、スコアリングの質管理、素点の確認、異なったフォーマットにおける点数の合算の正確さ、スコアや評価の正確性に関する質管理部分点の分析、合格判定の方法をスコアに当てはめる際の正確さ、学習者や教育者へのスコア報告に関する質管理、学習者にとってのスコアの記述や解釈の分かりやすさ、正確さ
内的構造	項目分析データ: ①項目の難易度と識別係数、②項目特性曲線、③項目間相関、④項目点と合計点との相関、項目の信頼性、測定の標準誤差、一般化可能性分析 <sup>[2]</sup> 、次元 (dimensionality) の分析、因子分析、差異項目機能、心理測定モデル
他の変数との関係	他の関連変数との相関、類似したテストとの収束的相関(内的、外的): 類似テスト、異なったテストとの弁別的相関(内的、外的): 関連のないテスト、テストと規準(クライテリア)との相関、エビデンスの一般化可能性 <sup>[3]</sup>
結果	テスト結果が学習者や社会に与える影響、学習者、今後の学習に対する影響、予測しなかった負の結果に対する正の結果の優位性、合格基準スコアの決定方法の妥当性、合格結果: ①決定の信頼性と分類の正確さ、②合格点における条件付き標準誤差、能力が不十分な合格者、能力が十分な不合格者、学習者に対する教育的影響

ことですが、これをやろうとすると途轍もなく労力が必要です。ただ、ひとつ重要なのは、ここが信頼性にあたる部分なので、信頼性というのは妥当性の一条件でしかない、ということが明記されていることです。

**21、Likert Scale:** ライカースケール、次の文章に対する最も近い印象を回答して下さい。大学において、教員は“教育学”に関して1年以上学ぶ時期があるべきである。こんな適当な項目を並べていますけど、こういうので1から5を選ぶというのがいわゆるライカースケールとよばれています。

**22、Likert法の件数の問題:** よく質問を頂くのですが、ライカート法で何件法にしたらよいのか、ここに5つ選択肢を設けてありますね。いろんな意見があります。数字が多い方が信頼性が高いと昔からよくいわれていますが、件数が多い方が統計学的に有意差を出しやすいということはいえます。それを言い出すとビジュアルアナログスケールという10cmの線があつて×をつけるようなやり方がありますけど、それが統計学的に有意差を出しやすいといわれています。でもこの一つ一つの数値がどのような意味合いをもっている

のか、というのをきちっと説明する anchor statement, descriptor とよばれていますけど、こういうのが数値が多くなると非常に困難になります。たとえばアメリカで Mini-CEX という診療能力の評価方法があります。プロフェッショナリズムを指導医が研修医に1～9までつけるという場合、7・8・9は上中下の上ですね。ここ(4～6)は中、ここ(1～3)は下。そのなかで、上の上は9をつけていけばいいわけですね。この評価をつけるのはそう難しくないという話もあるんですけど、1個1個に説明が付いている訳ではないです。この9件法というのはアメリカではけっこう受け入れられているものなんですけど、ヨーロッパではあまり見ません。奇数の件数については、5とか7とかですが、中心偏向性が問題になります。それから、ライカート法には、度合い0から10みたいな形のもの、肯定否定を含む-2から+2など0をまんなかにして0をはさむやり方もあります・・・。

**23、尺度変換:** さきほど100点満点のテストを間隔変数にしているかどうか、という話がありましたけど、尺度変換、順序変数を間隔変数に変えるというやり方があります。間隔変数は順序変数よりも統計学的な解析がしやすい、できればこれを1から5というのを間隔変数にしたいという衝動が起こるかもしれません。間隔変数に変換するということはよく行われていますし、平均点を計算することもよく行われています。PB

**Likert Scale**

- 次の方に対する最も近い印象を回答して下さい  
(1)全く同意、(2)あまり同意しない、(3)どちらでもない、  
(4)同意する、(5)全く同意する

大学において:

A) 教員は“教育学”に関して1年以上学ぶ時期があるべき

B) 教員はコミュニケーションスキルについて学ぶべき

C) 専門職の指導者は専門職意識を身につけておくべき

**Likert法の件数の問題**

- 何件法が良いかは様々な意見あり
- 件数多い方が差を際立たせやすいが、オプションの説明 (anchor statement, descriptor) 困難
  - 米国では9件法もよくみられるが、...

Professionalism (mini-CEX)								
1	2	3	4	5	6	7	8	9
Unsatisfactory			Satisfactory			Superior		

- 奇数件数は中心偏向性が問題とも言われる
- 度合(0～10など)と否定/肯定(-2～+2など)

**尺度変換**

- 間隔変数は順序変数よりも統計学的解析がしやすい
- 順序変数を間隔変数に変換すること、平均などを計算することもよく行われる

PBLを受けてどう感じたかについて回答して下さい。

5: 大変同意、4: 同意、3: 中間、2: 同意せず、1: 全く同意せず

a. 私はPBLに熱心に参加した

b. 私はPBLを面白いと思う

c. 私は学んだ内容に満足している

d. 私は後輩にPBLを勧めない

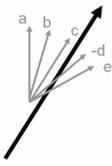
e. もしPBLと講義を選べるなら、私はPBLを選ぶ

Lについてどう感じたかについて回答して下さい。大変同意するから全く同意せず、まで、5から1で選びますと、1～5の数値に変換することも可能かもしれません。これらの項目の中でd、私は後輩にPBLを勧めない、がありますがこれだけは逆転項目です。

**24、内的一貫性:** a～e までの測定ベクトルの方向性が似ているなら、ひとまとめに合算することが許容される、dだけ完全に逆向いていますけど、当然逆転項目ですので、これは-dというふうにとってやるとかなりこちらの方向に向いているのがよくわかります。ですので、こういうふうにとりまとめにしてしまうということですね。このときにaからeの数値を合算してもいいかをみるためにこの内的一貫性というのを見るため

### 内的一貫性

- a～eの測定ベクトルの方向性が似ているならひとまとめに合算することが許容される
- もし複数の測定が互いにあまりに似ているなら、区別して測定する意味はない
- $0.7 < \alpha < 0.9$ が適正
- dは逆転項目



のクローンバックの alpha 係数が0.7から0.9だと大体同じ方向を向いているから合算していいだろうというかたちになります。dは逆転項目ですので、マイナスをつけてはじめて合算してよいということになります。Alpha が1.0だとどうなるか、もしほとんど同じ方向を向いている質問ばかりだと同じ内容が聞かれているわけですから、トートロジーになりますからそれはあまり意味がない、0.9以下の方が良いといわれています。

**25、尺度変換(続き):** もし、こういうふうな結果が出たとします、5・4・5・2・5という回答でしたら、合計得点は、 $5 + 4 + 5 + 4$  (これは逆転していますから $6 - 2 = 4$ )  $+ 5 = 23$ 、これがこの人に全体的な合算した得点です。内的な一貫性がなければ、スコアの合算は不可能になります。全てに5をつけているような学生がいるとすれば、逆転項目について読んでいないわけですから、こういうのは除いた方がいいということになります。

### 尺度変換(続き)

- 5 a. 私はPBLに熱心に参加した
- 4 b. 私はPBLを面白いと思う
- 5 c. 私は学んだ内容に満足している
- 2 d. 私は後輩にPBLを勧めない
- 5 e. もしPBLと講義が選べるなら、私はPBLを選ぶ

- 合計得点:  $5 + 4 + 5 + 4 + 5 = 23$  (dは逆転項目)
- もし、内的一貫性が低いならスコアの合算は不可能
- 全てに5を付けた場合、明らかに逆転項目は矛盾

**26、データ回収:** これが最後ですね。回収率が悪いデータというのはレスポンスのバイアスが生じる可能性が高い。臨床医の患者に対する態度を同僚評価したところ、回収率は20%で、9段階で平均7.5だった。7.5だったらいいじゃないかということになるのですが、態度に問題のある医師が同僚を評価せずに、あるいは同僚から評価されなかつただけかもしれません。回収率20%だったらそういう可能性は高いだ

### データ回収

- 回収率が悪いデータはレスポンスバイアス生じる可能性高まる
  - 臨床医の患者に対する態度を同僚評価したところ、回収率は20%だったが、9段階で平均7.5の結果
  - 態度に問題のある医師は同僚を評価せず、同僚から評価されないだけかもしれない
- 誰が回収するかによってデータの質は変化
  - 講義技法に問題あるという評判の教員が記名式の授業評価の用紙を配布・回収した
  - 結果は、回収率70%、5段階で平均4.2だった

ろうということですが。誰が回収するかによっても、データの質が変わります。例えば講義技法に問題があるといわれている教員が記名式の授業評価の用紙を自分で配布し、自分で回収した、という場合、回収率70%で、5段階で平均4.2だった。このような結果はどのくらい意味があるのかということとは微妙ですね。

いろいろな要因がありますけれども、教育現場で、特に問題のありそうなところをかいつまんでお話ししました。5分くらいオーバーしました。

Q: データ回収率は何%ぐらいないといけないのでしょうか?...

A: ものすごく回収率が悪いから、そのデータを出さないというのは問題だと思います。やったら一応回収率悪くても、こういう結果が出ました、こういう限界があります、と出して出す。回収率は70とか80%とかが安全域でしょうが。

Q: 逆転項目というのは入っていた方がいいのでしょうか?

A: そうですね。余り読まずにほとんど同じものにつける人がいるので。

Q: アメリカでは Likert スケールを何件法にする人が多いかについて、1995年には53241の順で今年の2005年には45321の順番になったと言われています。順番が変わった考えられる理由は?

A: それはリカート法の4件法が増えているということですか?(そうです、10年間で変わっているんです。)はっきりした背景はわかりませんが、5段階にしてその中心偏向性に大いに問題があったとか・詳しくはわかりません。

Q: アメリカではじまったステップⅡのCSも4件ですよ。

A: おそらく以前は4よりは5の方が信頼性が高くなるという、あるいはやっぱり5よりは6、6よりは7とどんどん増えていく方向にあったんですけど、信頼性自体が昔よりそんなに問題にされなくなってきたのではないかなとは思いますが。妥当性の方が大事だとはっきりとした定義されたことも1つの動きなのかもしれません。

Q: マクマスター大学の Norman さんは7件にということで教育しているということだったんですけど最近はどうではない?

A: あまりこだわる必要はないと思います。

Q: 逆転項目が矛盾している場合には?

A: それは取り除いて解析した方がいいでしょう。

Q: 心理学だと、出席回数は、比例尺度扱いすると思いますけど。

A: 比例も間隔尺度に含まれるとして、項目にあげなかったんです。

Q: 質問紙の作成に関してですが、スタンダードとなるような質問紙とかあるんでしょうか?

A: 質問紙だけを抜き出して本にしてあるものもあります。それを集めて紹介しているウェブサイトもあります。新しいものを作るのであればこのあたり、いくつか関連した本をここにおいてありますので休み時間にでも見てください。

## グループ討論1、まずやり方の説明:大西先生

次のプログラムのやり方の説明:シナリオ1というものがあります。シナリオがあった方が最初話し合いがスムーズに行くだろうということで設けています。これ以外のものがやりたいということであればそれ以外でも構いません。それからシナリオ1と2のどちらかだけでもいいですし、両方やってみようということであっても構いません。いずれにしても、シナリオ1, 2のほうですね何かデータをきちっと示して、次の教育改善のステップを進めようという流れになっています。そのためにどういった研究をするのか実際に計画をして、データを集めて、発表すればいいのかということを今の議論を踏まえて考えていただきたい、と思います。

### シナリオ1

あなたは、創立30周年を迎えた東京都立文京医療大学医学部の医学教育研究センター長に選出されました。先日学部長から早速相談を受け、その内容は次のようなものでした。「最近、臨床実習に出てからコミュニケーションが十分にとれない学生がいて困っている。英語、数学、理科(生物、化学、物理から2教科選択)、面接という入試の科目をどのように変更すればいいだろうか。特に、数学や理科については科目を減らしてもいいと考えているが、そのためにはこれを減らしても問題ないというエビデンスが欲しい。また、面接の方法を改善するか、新たな方法を加えるかして、これを研究面に結び付けてもらえると他の大学にとっても有用だろうと思っている。是非、色々調べて欲しい」

さて、あなたはどのような計画を立てますか。学部長はあなたに好意的に働きかけてくれています。また、文京医療大学の学生数、カリキュラム等については日本の一般的な像をイメージして、グループ内で決めていただいで結構です。

### シナリオ2

あなたは、300床の区立病院総合診療部部長であり、昨年から2週毎に「プライマリケア研修」と称して文京医療大学医学部の学生を2名ずつ研修に受け入れています。外来で実際の患者と面接する実習を行い、看護師たちには「医学生が面接をしても、今までのところ患者さんから不満は出ていないし、大丈夫だと思っているんだ。それと、2週目にはとても彼らは伸びるから、教育効果はとても高いよ」と説明しています。しかし、このことを客観的に表すデータは今のところ持ち合わせていません。

外来では医師、看護師共に、診療時間中に学生を評価するような時間はありません。ただ、外来医師は学生が面接し、カルテを簡単にまとめた後で学生と1分ほどディスカッションする「1分間プリセプティング法」によって指導することにしています。このような指導体制に問題がないこと、教育効果があがっていることをどのように示せばよいでしょうか。

(それでは、グループで移動されて、プロダクトをもってまたここに集まってください。次の集合時間は1600に集合してください。1600から全体討論をします。発表者の方はその準備もして下さい。6分間[発表4分、質疑応答2分間]とします。

## グループ討論1:シナリオに沿った研究計画—デザイン、サンプリング

### 全体発表と討論

**1班:**シナリオ1を選択しました。デザインの評価をする前に、コミュニケーション能力をどのように

評価するののかということをもまず学力に関係しているのかということに関して、構成概念化が必要だったわけですが、それについてかなり時間を費やしてしまいました。デザインをはっきりと明示することはできませんでしたが、簡単に発表させていただきます。まず、試験改革の目標を設定させてもらいました。いろいろパターンがあると思いましたが、

(1) 試験改革の目標

- ①はじめから、コミュニケーション能力の高い学生をとる……採用
- ②コミュニケーション能力を高め易い学生をとる ……不採用
- ③コミュニケーション能力が低い場合退出してもらいやすい…不採用

(1) 初めからコミュニケーション能力の高い学生をとることをシナリオ1で目標とすることを仮定して、デザインを進めてみました。

(2) したらよいかも知れないこと(これらの改革の価値を見通す研究をする)

- ① 学科試験をへらす 例: 5科目→4科目、時間削減。
- ② 面接の内容を変更する 例: 集団面接を加える
- ③ 他の方法を加える 例: ポートフォリオ

(2) どのようなことをしたらよいのかですね、すなわち教育的な介入ですけれども、学科試験を1つ減らすということが考えられました。それから面接の内容を変更する。たとえば集団面接を加えてコミュニケーション能力を測ってみるといふこと、

(3) 研究のデザイン

- ① 長期的には、4年後の結果で最終的な検証を行う(共用試験or大学オリジナル)
- ② 変数関係のモデルをつくる  
コミュニケーション能力というものの構造を明確化するために、調査が必要。そこをまず調査して、さらに入試との関係をモデル化して、変数関係のモデルをつくる。そのシミュレーション結果をもとに、入試をつくっていく。
- ③ どのような問題が起きているか、把握する&シミュレーションも必要  
理系科目を減らすべきではないかもしれない。バランスがよい方がコミュニケーション能力がよくなるかもしれない。受験科目を減らすと、受験勉強の仕方も変わる。

それで(3) 研究のデザインですが、①4年後の結果: コミュニケーション能力のある学生を1年目に入れたときに4年後どうなっているのかをみるためには少なくとも4年、5年はかかりますので、待てるかという話しになりまして、とりあえず、4年後の結果と現状を見る、の2つ…というディスカッションになりました。1つはですね、4年後どのような評価をするのかということで、全国一斉の共用試験がありまして、共用試験におきましてまず学力を測定できたらどうかと考えられました。それからコミュニケーション能力が臨床実習に出る前に十分な能力を有しているかということについて

はOSCEの医療面接だけでは測りきれないだろうとして、評価方法については今検討中で、大学オリジナルの方法で4年後評価するという事です。②変数関係のモデルをつくる③どういう問題が起きているのかを把握する、

(4)試験の変化の方向

①受験生同士のコミュニケーション、をチェックする

(試験官と受験生だけのコミュニケーションではだめ)

②一般科目の能力と、コミュニケーション能力の関係を検討する

(5)学生の能力の評価方法

調査紙法(学生同士、患者から、その他の人から)

(4)試験の変化の方向、その後いろいろディスカッションがあったのですが。全体をまとめますと、入試の制度を少し変える、学科を減らしたことで4年後の学力が落ちていないこと、コミュニケーション能力がしっかりしていることをみるような研究のデザインになればいいかなと思います…。これは…コミュニケーション能力に関する構成概念について、ディスカッションが続いていたものですから、…一般科目の能力とコミュニケーションの関係を検討するようになりサーチにしたいというのがひとつ話題に上がっております。(5)コミュニケーション能力の評価方法:コミュニケーションを評価することに質問表を作成したり、質問紙評価は学生同士、患者様から、教官の評価をうまくパッケージ化して評価を行う方法としたい。以上です。

追加:コミュニケーションの構造「何がコミュニケーションなのか」というのを調べるのには、臨床実習なんかに関係した、実際にそこで起きている問題を調べていった方がいいのでは、③のところですね。それが分からないと、測定もできないし、対策もたてられないということです。

Q:これは要するに臨床実習に5年なら5年生が実習のときに数学理科の成績と面接とどれくらい相関があるかを調べるのですか

A:相関があるだろうという仮説からです。

Q:ポートフォリオとは具体的にどういうことですか

A:まだ十分話しをしていません。高校時代の活動を面接試験の判定に活用する、2年間の医学部における成長を判断するときに活用する、などを考えています。評価の一部として考えました。

C:評価は難しいですね。

**2班:**最初に手持ちのデータを使った…入試成績のデータ…現在の成績というのがあります。成績には2種類あって、1つが知識・技術を評価するような成績と、もう一つがコミュニケーション能力とか問題点とかいうこれをどう評価するのか議論はあるのですが、いずれにせよこういう質的なデータを何らかの形で量的なデータに変換しまして、入試成績と現在の成績との関係を検討しま

す。検討の仕方は統計学的方法で相関・回帰分析とか・・・いずれにしても段階目は入試成績と

<p>コミュニケーション能力とは何か</p> <ul style="list-style-type: none"> <li>質的研究などで要素の抽出</li> <li>面接試験のチェックポイントとの照らし合わせ</li> </ul>	<p>デザインの大枠</p> <ul style="list-style-type: none"> <li>パイロットで事後テストのみ???</li> </ul>	<p>サンプリング</p> <ul style="list-style-type: none"> <li>前後期入試で分ける</li> <li>推薦入試と一般入試</li> <li>5年生にパイロット研究をして改善点を確認する。 新しいコミュニケーション能力の試験をして手もちのデータと比較する。</li> </ul>
--	---	---

<p>問題点</p> <ul style="list-style-type: none"> <li>入試でコミュニケーションを測れるのか？ → 測れることとする</li> <li>科目を減らすのか</li> <li>減らした群と減らしていない群との比較にするのか？</li> <li>コミュニケーション能力と科目との相関は？</li> <li>Control study をしているのか？</li> <li>面接はコミュニケーション能力を予測しているのか？</li> </ul>	<p>Pilot Study (retrospective) 1</p> <ul style="list-style-type: none"> <li>入試の得点(数学・理科・英語)とコミュニケーション能力の相関を見る → 入試科目とコミュニケーション能力と現在の成績を見て、コミュニケーション能力の低い学生が平均点の低い入試科目の配点を下げる(段階的に) → 臨床実習に来ている5年生の学生のコミュニケーション能力を測る             <ul style="list-style-type: none"> <li>5年前の入試の成績との相関を見る</li> <li>コミュニケーション能力の測定方法はここでは問わない</li> </ul> </li> <li>現在の5年生のコミュニケーション能力の弱点を知る</li> </ul>	<p>Cohort Studyのデザイン</p> <ol style="list-style-type: none"> <li>面接でコミュニケーション能力をより正確に測定する方法を考案する 面接をOSCEのような設定にする 小論文をアレンジする (面接の目的は？見切りではないもの)</li> <li>従来の方法と新しい方法の面接で学生を選抜する 全員に2つの面接方法を行い、選抜するときに2群に分ける 入試科目を1科目減らす 入試科目とコミュニケーション能力と現在の成績を見て、コミュニケーション能力の低い学生が平均点の低い入試科目の配点を下げる(段階的に)</li> <li>毎年コミュニケーション能力を測定する</li> <li>コミュニケーション能力と成績との両方を評価する(5年後)</li> </ol>
---	---	--

現在との関係を見るということをや。今やることというのは学年に応じたコミュニケーション能力の成績ということですが、コミュニケーションをどう評価するか・・・できるということにしておきます。例えばオスキーや他の方法を使ってもできるとします。科目の削減や1段階目の分析結果を得て、ある科目はなくすだとか、というのをやります。サンプリングこれはデザイン

<p>問題点</p> <ul style="list-style-type: none"> <li>入試でコミュニケーションを測れるのか？</li> <li>科目を減らすのか</li> <li>減らした群と減らしていない群との比較にするのか？</li> <li>コミュニケーション能力と科目との相関は？</li> <li>Control study をしているのか？</li> <li>面接はコミュニケーション能力を予測しているのか？</li> </ul>
--

と関係していますが、前期は従来型で後期の方を変更型にして比較する、もしくは推薦と改良型の比較検討する。デザインとしては準実験的デザインになると思います。この場合ランダム化していないので質の保障は得られない。という大きな問題があります。5年生のパイロット研究保留した新しいコミュニケーション評価の方法と面接方法との相関があるのか、確かめておかなければならない。デザインの大枠としてパイロットでは事後テストのみになるが全体としては新しいやり方を確かめるには準実験的なものになるだろう。以上です。

Q: 科目を減らすのは相関がなければ、減らしていいのでしょうか？科目数の増減に相関は？余り関係ないでしょうか？

Q: 新しい体制にするというとき、今あるデータは余り根拠にならない？

A: 次のデータを見るときにはもっといいデザインにしていく方がいいでしょう。

**3班**: 2班とほぼ同様な形になっています。最初ブレインストーミングで問題点について話しました。コミュニケーション能力は測定可能なのか、科目を減らして本当に大丈夫なのか、減らした群と減らさない群の比較をするということは、かなりリスクではないかという意見も出ました。コミュニケーション能力と入試科目は相関するかどうかを見るべきでしょう。入試科目を減らした群とコントロール群としていいのか倫理的な問題、リスクの問題を考えるといいのか、コミュニケーション能力の測定を5年後しても意味があるのか、ということが話題になりました。デザインとして大学のイメージとして一般的に学生100名、従来の面接は一次試験で学力判定をして定員の3分の1位にとって、二次試験で面接を行っていたと、面接も5段階評価で、教官2名で志望動機をきくなど良くあるパターン、問題ありの学生を切っていく形の入試のあり方、を行っていたという想定をしました。今あるデータで、パイロットスタディをみてることにしました。入試科目の数学理科英語コミュニケーション能力の関係をみていく。具体的には臨床実習にきている5年生のコミュニケーション能力を測る、量的に測定可能ということで進める。その学生の5年前の入試の成績との相関をみることにしました。ここでエビデンスを得ることにしました。また、現在の学生のコミュニケーション能力の弱点具体的にどんなことが問題なのか、を知ろうということにしました。次に従来の面接方法の問題点はどうであるか、評価者のばらつきがあったのか、そういう点をきちんと評価しよう、合格判定において面接がどれくらいの比重を占めていたのか、も考慮。面接でコミュニケーション能力を正確に判定できるような方法を考案する。もう少しアクティブにコミュニケーション能力を測定できる方法を考える、ことにしました。コントロール群にしてしまうと後で評価ができなくなってしまうので全員に2つの方法を実施した上でコントロール群を設定する。毎年コミュニケーション能力を測定する。5年後成績とコミュニケーション能力を測る、但し成績とコミュニケーション能力のバランスが大切であることは基本です。

C:入試科目に国語を入れるべきだという意見もありました。

### デザインとサンプリング

- 大学のイメージ
  - 1学年100名
  - 従来の面接は、「一次試験で学力判定を行い、二次試験で面接を行なう」という方式

### Pilot Study (retrospective) 1

- 入試の得点(数学・理科・英語)とコミュニケーション能力の相関を見る
  - 臨床実習にきている5年生の学生のコミュニケーション能力を測る
    - 5年前の入試の成績との相関を見る
    - コミュニケーション能力の測定方法はここでは問わない
- 現在の5年生のコミュニケーション能力の弱点を知る

### Pilot Study (retrospective) 2

- 従来の面接方法の評価と問題点の抽出
- 合格判定における面接の評価の比重は？

### Cohort Studyのデザイン

1. 面接でコミュニケーション能力をより正確に測定する方法を考案する
2. 面接をOSCEのような設定にする  
小論文をアレンジする  
(面接の目的は？ 足切りではないもの)
3. 従来の方法と新しい方法の面接で学生を選抜する  
全員に2つの面接方法を行い、選抜するときに2群に分ける
4. 毎年コミュニケーション能力を測定する
5. コミュニケーション能力と成績との両方を評価する(5年後)

Q:最近とはどういう意味ですか？

A:要するに臨床実習が参加型になってその部分がクローズアップされたという印象で書きました。

Q:コミュニケーションに問題ありとは、どういうレベルですか？例えば患者さんに話ができないとか…挨拶もできないのか。

A:全てのレベルです。今そういう人を排除する方法がないことが問題になっていると思います。

Q:大学によって5年生に求める能力にはかなり差があるように感じています。一応臨床参加型を想定すればいいですか？

A:そういうことになります。そのときに問題にならないように。今OSCEや共用試験が入ってきているという背景から非常に重要なことだと思います。

Q:選択科目がちがうのでその場合どう評価するのかという点、このシナリオだと面接でコミュニケーション能力のことは意識外のことだったということでランダム化されているのでは、という点

A:能力のバランスを重視した。選択科目に関しては具体的なことは十分議論されていません。

Q:旧面接試験と新面接試験を5年かけてどっちがいいんだみたいな、その前の段階で面接の違いと言うのをもう少し絞るような形でこれをやった方がいいのでは、

A:確かに5年待つのはつらいという話はしました。毎年測定して次の学年にいかしていくのかなというイメージはあります。単純比較ではないかもしれませんが。介入の有無については話題になりました。

次は1650からにします。休憩です。再集合は1735です

## グループ討論2:シナリオに沿った研究計画—データと解析の種類

### 全体発表と討論

**3班:** 1)コミュニケーションを取れる学生を入試でとるには、入試科目とコミュニケーション能力と現在の成績をみて、コミュニケーション能力の低い学生が平均点の低い入試科目を減らす。2)新たな疑問:①5年次のコミュニケーション能力が悪ければだめなのか—教育効果は期待できるのか。②コミュニケーション能力をどう評価するのか 説明者と被説明者との理解のズレ、またこの学生に診て欲しいですか 3)入試の新しい方法 ①上級生が学生のグループ面接に入り評価する ②天才系は落ちる ③1回の試験では評価できない ④

#### 新たな疑問

- 5年次のコミュニケーション能力が悪ければだめなのか
  - 教育効果は期待できないのか
- 毎年評価することでフィードバックをしていく
- コミュニケーション能力をどう評価するのか

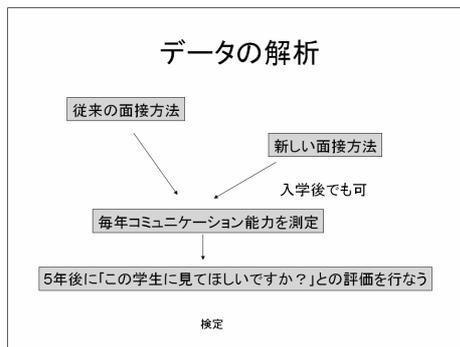
#### コミュニケーション能力をどう評価するか

- 説明者と被説明者との理解のずれを評価
- 360度評価(教員・学生同士・コ・メディカルなど)
- 「また、この学生に診てほしいですか？」
- コミュニケーション能力と臨床能力との相関は？

集団生活をさせて行動を判定する ⑤想定不可能なシナリオを作り反応を見る。データの解析は、従来の面接方法と新しい面接方法

入試時のコミュニケーション能力測定法は？

- ・ 上級生が学生のグループ面接に入り評価する
- ・ 天才系は落ちてしまう？
- ・ 1回の測定では評価できない
- ・ 集団生活をさせて行動を判定する
- ・ 想定不可能なシナリオをつくり、反応をみる
- ・ クレペリンテストなどを利用する



毎年コミュニケーション能力を測定する→5年後に「この学生に診てもらいたいですか」との評価を行う→ $\chi^2$  検定を行う。

Q:誰が評価しますか

**1班:** 1)コミュニケーション能力の現状調査—形式(観察・チェックリストの有無等)、コミュニケーション能力と入試科目の相関 2)コミュニケーション能力とは、患者の心理的状況を把握できる・トラブルからの対応ができる・個人の能力・社会での能力・医療での場面における適応できる判定法・問題点は何か・観察が重要・つまづき 3)コミュニケーションの問題点? どういうデータを収集すべきか・形式 観察が重要 質問紙法 観察チェックリスト 対象:学生 項目 誰が測る 学生自身 学生同士 面接:尺度をどうする 誰が、項目 測る人にとって差[+] データ:現場で何が起きているか 問題を起こした学生のエピソード[聞き取り]質的に基づいてモデル項目リストの作成 解析 全体的にコミュニケーションレベルの低いのを入れない、伸びない人をとるのか?

<ul style="list-style-type: none"> <li>・ コミュニケーション能力の現状調査 <ul style="list-style-type: none"> <li>- 形式(観察チェックリストの有無 等)</li> </ul> </li> <li>・ コミュニケーション能力と入試科目との相関 <ul style="list-style-type: none"> <li>- 入試科目を数や種類の検討</li> </ul> </li> </ul>	<p>コミュニケーション能力とは</p> <ul style="list-style-type: none"> <li>・ 患者さんの心理社会的状況を把握できること?</li> <li>・ トラブルからの対応ができるか</li> <li>・ 個人の能力か?</li> <li>・ 社会での能力か?</li> <li>・ 医療での場面における適応できる判定法</li> <li>・ 問題点は何にか?</li> <li>・ 観察 <ul style="list-style-type: none"> <li>どういった場面でつまづいているか</li> </ul> </li> </ul>	<p>コミュニケーション能力の問題点は何にか?</p>
<p>どういったデータを収集すべきか</p> <ul style="list-style-type: none"> <li>・ 形式 観察が重要! <ul style="list-style-type: none"> <li>- 質問紙法</li> <li>- 観察 チェックリスト</li> </ul> </li> <li>・ 対象 <ul style="list-style-type: none"> <li>- 学生 <ul style="list-style-type: none"> <li>コミュニケーションがつかないかを事前に把握</li> <li>(学年間、臨床実習に出ているか否か、臨床実習&lt;開始と1年後&gt;の時期)</li> </ul> </li> </ul> </li> <li>・ 項目 <ul style="list-style-type: none"> <li>- 指導の問題は?(伸びるかどうか)</li> <li>- 多すぎるファクターへの対応</li> </ul> </li> <li>・ 誰が計る そくいてい <ul style="list-style-type: none"> <li>- 学生自身 学生同士</li> <li>- 患者さん</li> <li>- スタッフ 指導医</li> </ul> </li> <li>・ どうやって</li> </ul>	<p>観察</p> <ul style="list-style-type: none"> <li>・ 尺度をどうするか</li> <li>・ 誰が</li> <li>・ 項目 <ul style="list-style-type: none"> <li>- 学生自身 学生同士・患者さん・スタッフ から</li> <li>- 計る人によって項目が異なるか? 異なる</li> </ul> </li> </ul>	<p>臨床実習におけるコミュニケーション能力を測定するには</p> <ul style="list-style-type: none"> <li>・ データ <ul style="list-style-type: none"> <li>- コミュニケーション能力に問題が起きているとは?</li> <li>- 現場では何が起きているか? 問題を起こした学生のエピソード (聞き取り)質的</li> <li>- 基づいて モデル項目リストの作成</li> </ul> </li> <li>・ 解析</li> </ul>

**2班:**入試の改善には 目的:コミュニケーション能力の高い学生を選抜 期間:5 年間 プレ・ポス

<p style="text-align: center;"><b>アイデア</b></p> <ul style="list-style-type: none"> <li>・前後期入試で分ける</li> <li>・推薦入試と一般入試</li> </ul>	<p style="text-align: center;"><b>アイデア</b></p> <ul style="list-style-type: none"> <li>・<b>提案Ⅰ</b> 前期⇒従来型 後期⇒改良・新型 ※配点については非公開</li> <li>・<b>提案Ⅱ</b> 前期 後期⇒ 従来型と新型/改良型両方 ※100人の受験生を対象に ※選抜基準が異なる、配点が異なる⇒研究として行ってみる ※評価方法(新・旧)をランダムに分ける</li> </ul>	<p style="text-align: center;"><b>データ</b></p> <p>前期/後期⇒ 従来型と改良型両方を行う ※100人の受験生を対象に ※選抜基準が異なる、配点が異なる⇒研究として行ってみる ※評価方法(新・旧)をランダムに分ける</p> <ul style="list-style-type: none"> <li>・コミュニケーションに問題のある学生には介入⇒介入の定義⇒改善されたか</li> <li>・入試で コミュニケーション能力の高い学生を選抜することが目的</li> </ul>
---	---	---

トの比較 サンプルング:まず1年生対象[+5年生でも行う]  
方法:前半を従来型、後半を新型、学生の質が違い[+]→コントロール群になりえない  
コミュニケーションの評価:どのように審査する→同僚評価は進級対象にしない、チュートリアルを評価する、教員による面接評価、グループディスカッションをさせたらどうか、評価者のレベルを上げる、短期2年間、長期5年間の変化を観察する

Q:同僚評価の信頼性?

Q:全面改訂するリスク?

<p style="text-align: center;"><b>入試の改善には</b></p> <ul style="list-style-type: none"> <li>・入試でコミュニケーション能力の高い学生を選抜することが目的</li> <li>・五年間の調査(事前/事後の比較)</li> <li>・サンプルング・・・一年生を対象(+五年生にも行う)</li> <li>・入試変更前・・・現一年終了時に評価 新入一年生終了時に評価 学年ごとに別の評価方法⇒二学年を比較</li> </ul>	<p style="text-align: center;"><b>解析の種類</b></p> <ul style="list-style-type: none"> <li>・二群の変化</li> <li>・コホート調査(個人の変化追跡調査?)</li> <li>・面接と学科成績の相関</li> </ul> <p>⇒全体のレベルupなのか、個人の能力upの問題なのか? ⇒コミュニケーション能力とは?</p>
--	---

<p style="text-align: center;"><b>目的</b></p> <ul style="list-style-type: none"> <li>・学部長の意図・・・入試時にコミュニケーションに問題のある学生を選別</li> <li>・全体レベルupの改善 ⇒データ分析しやすい 全体のレベルの改善により、個々の問題も改善してくるのでは?</li> <li>・短期計画⇒二年間の前後比較 現状から判断して入試方法を改善</li> <li>・長期計画⇒五年間の変化を観察</li> </ul>	<p style="text-align: center;"><b>コミュニケーション能力の評価</b></p> <ul style="list-style-type: none"> <li>・コミュニケーション能力をどのように審査するのか ⇒短時間の面接では評価しにくいのでは?</li> <li>・同僚評価⇒進級評価などには利用可能</li> <li>・チュートリアル評価</li> <li>・教員による面接評価⇒この改善? (教員が特定の学生を継続的な面接評価をするなどの工夫が必要)</li> <li>・グループディスカッションを観察</li> <li>・評価者の問題</li> </ul>
--	---

本日最後の講演になります。医学教育研究の現状と課題ということで・・

## 講演：医学教育研究の現状と課題

### 東京大学医学教育国際協力研究センター 大西弘高先生

現状と課題ということですが、・・・今わかっている内容を話していきます。また中程で医学教育の現状を探ってみようということで、医学教育学会で出している緑色の雑誌ですね、あれの5年間出ている論文をこちらのセンターに研究生としておられた足立先生が分析してくれています。彼は今ダンディー大学スコットランドに行っているんですけど、もうすぐスコットランドからみえるはずなんです。雪で飛行機が飛ばなかったとかで、さっき成田についたという連絡があったらしいので、なんとか間に合ってくれると思います。

**1、なぜ医学教育研究が必要か**：まず、そもそもなぜ医学教育研究が必要かというところにもう一度立ち返ってみます。教育ということに関して考えると必ず誰もが経験しているので、語りたがっている人はいっぱいいるんですけど、その議論の多くは経験的なレベルに留まってしまうということがあります。ただ、経験的なレベルにとどまっても声の大きい人が勝つということもあります。各医学専門分野が最近「教育」のいろいろな業務を

切り離して研究に専念したいという想いが強くてそれを組織的に在る程度後押ししようというふうな感じもしてきています。それがおそらく教育専門の部門等をたくさん作っていることに現れているのではないかと思います。教育専門の部門になりますと、その人の業績、教育そのものを業績にしていけないといけない、ただそれが教育専門の人だけに当てはまるのかもしれないし、それ以外の人にも当てはまるのかもしれない。とにかく教育を業績としてどのように測定したらいいのかということでおそらく研究をして論文を書きなさいとかそういうふうなプレッシャーが随分強まっているような印象を受けます。これは全世界的にこうふうな印象だといえます。

**2、医学教育研究英語論文('94-'03)**：医学教育研究の英語論文というのが10年間くらいで、どのくらいあるかということと24000編ほどあります。これは多いなと思いました。Academic Medicine とか Medical Education が・・・2つのこれでたった17%しか占めていないこの数字にもびっくりしました。Family Medicine というのは Family Practice のアメリカの学会の雑誌です。これは教育のみを扱う雑誌です。他に Practice 自身を扱って

いるものもあります。あとBritish Medical journal、今医学教育に力をいれていますが非常に多いです。あとJAMA とか CMAJ(Canadian Medical Association Journal)があります。こういったところも結構取り上げています。日本医師会では医学教育の特集を見たことがないんですけど。

#### なぜ医学教育研究が必要か

- 教育は誰もが経験し語りたくなるものだが、その議論の多くは“経験的”なレベル
- 各医学専門分野が「教育」を切り離し研究に専念したいという想いを持っている？
- 教育専門部門では「教育そのものを業績とする」流れが重視される

#### 医学教育研究英語論文('94-'03)

- 全体で24,028編
- Academic Medicine 11.2%
- Medical Education 6.0%
- Family Medicine 3.0%
- British Medical Journal 2.7%
- JAMA 1.5%
- CMAJ 1.3%

(Wartman. Acad Med 2004; 79: 910-17)

**3、研究論文内容:** イリノイ大学で同級だった人がこういう研究をやっていました。Academic Medicine, Medical Education, Teaching and Learning in Medicine のメジャー3誌においていろんな内容を吟味していった。その結果、その研究の内容というのは医療者の質とか患者のアウトカムとか肝心なことにはあまり触れていない。先も言った様に、いろんな評価のレベルがありますから、患者のアウトカムというのが一番高いレベルですが、そこまではなかなかいかないということです。学習者の評価データ、あるいは学習者の満足度というのが主な内容になっているに過ぎない。これは非常に厳しい意見ではありますけど、やはりコミュニケーション能力などという複合的なところはしっかり測定しようというのはなかなか難しい、結局はあまりできていない、ということがわかります。

**研究論文内容**

- Acad Med, Med Teach, Teach Learn Medのメジャー3誌において
- 内容は医療者の質や患者のアウトカムといった肝心なことには触れていない
- 学習者の評価データ、学習者の満足度が主な内容となっているに過ぎない  
(Prystowsky & Bordage, Med Educ 2001; 35: 331-6)

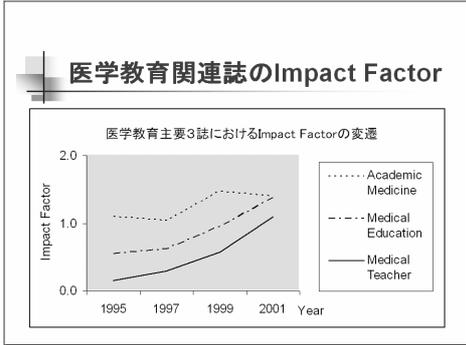
**4、Glenn Regehr の言う主要4誌:** Academic Medicine の去年の10月号でしたか、メディカル・エデュケーション・リサーチの特集をやっていました。そのなかで Glenn Regehr というのはトロント大学で外科部門があり、OSCE をカナダで開発してきているので有名なのですが、その外科の医学教育部門の中の心理学者ですね、外科の中に心理学者を雇っている、おもしろい部門ですが、彼が言うにはこの4誌が今のところ主要ではないか

**Glenn Regehrの言う主要4誌**

- Academic Medicine
- Advances in Health Sciences Education
- Medical Education
- Teaching and Learning in Medicine  
(Regehr G. Trends in Medical Education Research. Acad Med 2004; 79: 939-47)

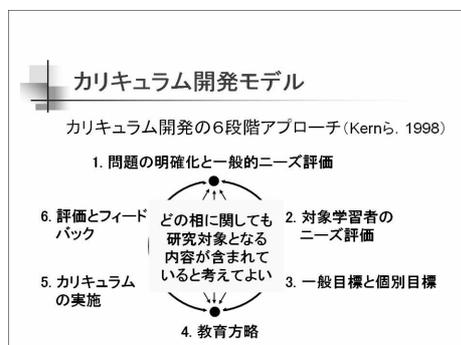
といっています。Advances in Health Sciences Education というのはマックマスターの Geoff Norman が中心に出している雑誌です。今非常に定評は上っています。ただ歴史がないのでインパクトファクター自身はあがっていません。Teaching and Learning in Medicine は Barrows がずいぶんがんばってきたようですが、最近の記事がないようです。

**5、医学教育関連誌の Impact Factor:** インパクトファクターはどうなっているんだろうかと思って、結構昔からデータを集めていまして、まとめていました。2001年では Academic Medicine と Medical Education が非常に拮抗しています。今 Medical Education が抜いています。Academic Medicine はアメリカの医学部長病院長会議がやっていますけど、アメリカのローカルな話題が多すぎて、アメリカがグローバルと考えられているらしいですがそのことが世界にはあまり受け入れられていないようです。Medical Teacher というのは足立先生が行っているダンディー大学の Harden 先生が中心に編集をやっていらっしゃるものです。今急激に伸びています。ただレビュー記事が多いという批判はあります。

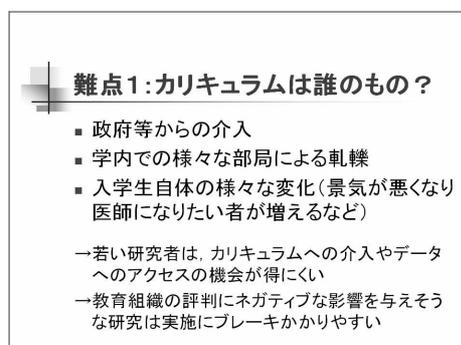


**6、カリキュラム開発モデル:** 先程のカリキュラムが研究の対象になることが多いだろうという話で

すが、実際にはカリキュラム全体ではなくても、それぞれの学生とあるいは患者さんとが学び合っているなかの状況とか環境とかいろんなくらいでも研究にするネタはあります。ただカリキュラムだけをこういう風に捉えていっても、例えばこのニーズ評価だけを取り出して調べるというふうなことでもいいわけですね。例えば先程のシナリオでも今の面接方法にどのような問題があるのか、コミュニケーション能力って何だ、どのように測るんだ、というのはこのあたりですから、そのようなことだけでも見直ししていくくらいでもテーマはたくさんあります。どのようなフェーズにしても、研究対象になるようなものはあります。特に最近このカリキュラムの実施は、どのような組織体系にしていると、うまくことが運ぶのか、そういったことも研究対象になりえます。



**7、難点1:カリキュラムは誰のもの?**:これは去年の Academic Medicine の中の特集のなかのもので、なぜ医学教育研究は難しいのかという論文です。その中でいろいろ出てきたのですが、これというのを抜き出してきました。まず難しいのはまず、カリキュラムは誰のものなのか? 政府等からの介入がある。アメリカだと政府よりも教育組織ですが、日本は本当に政府中心に医学教育改革が進んで、社会主義国家といえます。学内と



ということだけでも様々な部局による軋轢が起こりえると思います。例えば医学部長がこういうふうになれと言ったので、今こういうふうな研究をやりたいということ、そういうシナリオになっていきますよね。でも5年くらいの長いスパンを皆さん考えておられましたけど、5年間医学部長が変わらないというのは、あまりないですよ。そこで5年間のスタディというのは非常にリスクだと、3年後くらいに学部長が変わって、梯子を外されてということが起こりうる。それから入学生自身もどんどん変化していきます。例えば景気が悪くなって医師になりたい者が増えて私立でも一番下の偏差値というのがこの20年くらい上がってきたんですよ。そういったことも踏まえて、いろんな変化についていかなければいけないがあります。さらに難しいのは若い研究者が医学教育に興味をもっていろいろ勉強する、このカリキュラムをこう変えようとするのではないかといっぱいいろんなアイデアを持つんですけど、カリキュラムをこういうふうにしたらどうですかと上の先生に言ったら、それは10年早いとかそういう風に日本では言われがちですよ、それから入試データが欲しいというのどきたい金庫の奥のほうに入っていますよね、そういうのをうまく取り出すことが難しいかもしれません。それからいろいろとごちゃごちゃとやりはじめると、医学教育のネガティブな結果が出てくる、するとその結果を非常に上の人たちが気にする、そういうことが実際にブレーキになりやすく非常に難しいですね。

**8、難点2:概念的な問題:**それから概念的な点ですね、交絡因子が多い、いろんな背景になる要

困ですね、真の要因というのが把握しにくい。特に今の  
ような何を測定すればいいのか分かりにくい場合には、  
いろんな要因を測ってとりあえず全部統計的な分析に  
入れてしまえ、あるいは質的にいろんな概念を入れて、  
とりあえず何かを生み出そうというふうになっています。  
結局、統計学的な解析を使うにしても非常に複雑さが  
増すというのは大体心理教育系の特徴です。それから  
学生の動機づけとか、指導者のスキルなどというものは  
あまりシステムにのってこないものなんですけど、こういった複雑な要因も関係しています。さら  
に信頼性のある測定が難しい、ということもあります。

**9、難点3: 経済的理由:** 経済的な理由ですね、大体科  
研費のシーズンが終わって今度は厚生科研費の締め  
切りが、という季節になっていますが、研究補助金とい  
うのが教育研究にはつきにくい印象があります。文部  
科学省の科学研究費も医学教育研究というそういうカ  
テゴリーそのものがないですねまだ、それから例えば企  
業とかがお金を出してくれる、お金自体はそんなに沢山  
無くても教育研究は割りとできると思うのですが、製  
薬会社に生物医学的なものを持って行ってというのはわりとできますが、教育研究の場合なか  
か難しい。京都科学やレールダルにせいぜいもって行ってお願いするくらいかもしれませんね。

**10、難点4: アウトカムとの関連性:** さらにアウトカムで  
すね、例えばFDの効果を問うような研究をしようとする  
なら、教員の教育技法が向上し、それによって学習者  
の能力が向上して、患者のアウトカムが改善されてとい  
う3段重ねくらいでようやく影響が出てくるわけです。例  
えばPBLが生涯学習に正の効果をもたらすかどうかと  
いうのは10年ぐらいたってはじめてようやく判断可能と  
いうことがあります。そのときに例えば3年計画ぐらいで

研究計画を書けといわれても非常に難しいというのが現状だろうと思います。結局、どのような代  
理になるアウトカムによって短期間に結果を見ることが可能かどうか、こういうのを探ることが可能  
かどうか今回のシナリオでもあったように重要で難しい点ではないかというふうな印象があります。  
また、出資団体はですね、非常に短期間にアウトカムの報告を要求する習慣になりつつあります。  
この方向は日本でも強まっています。

じゃ、ここから足立先生お願いします。

#### 難点2: 概念的な問題

- 交絡因子が多く、真の要因が把握しにくい
  - 必要となる統計学的解析も生物学的研究に比べて複雑さが増す
- 学生の動機づけ、指導者のスキルなどの複雑な要因が関係する
- 要因のいくつかは、信頼性のある測定が難しい

#### 難点3: 経済的理由

- 研究補助金は教育研究にはつきにくい
  - 科学研究費も“医学教育研究”という範疇そのものがない
- 研究結果を企業等のインセンティブに関連づけることが非常に困難

#### 難点4: アウトカムとの関連性

- FDの効果を問う研究なら、教員の教育技法向上→学習者能力向上→患者アウトカム改善という形で影響が及ぶ
- PBLが生涯学習に正の効果をもたらすかどうかは10年ぐら이의スパンでようやく判断可能
- どのような代理アウトカム (surrogate outcome) によって短期に結果を見ることが可能か
  - 出資団体が短期的アウトカムの報告を要求する習慣

遅くなってすみません。足立と申します。今ダンディー大学医学教育センターに大学院生として留学しています。昨日雪のエジンバラ空港で足止めをくってしまいました。到着するのが遅くなってしまいました・・・。

**11、医学教育研究における研究手法の分析 —最近の傾向—**

**12、医学教育研究の特徴:** 医学教育研究の特徴としましては3つあげましたが、研究デザインはなかなかこれは何が一番いいかというのは定義によっても難しく、研究者の工夫によるところが大きい。それから研究費は一部の教材開発ですとかにつき、あと非常に少額のことが多いです。それから3点目として、臨床研究でなされているような大規模研究、ランダム化、盲検化された研究は一般的に困難であるということがいわれます。

**医学教育研究の特徴**

- 研究デザインは研究者の工夫によるところが大きい
- 研究費は少額のことが多い
- 大規模研究、ランダム化、盲検化は一般に困難

**13、対象・方法:** 今回の研究の目的としましては、こういった制約がありながら、どういった研究方法が可能かということを探るということでありました。対象・方法ですが、医学教育雑誌に掲載されました論文2001年の2月、第1号から最新号2005年10月まで掲載された論文を取り込み基準としました。除外対象としては、データの数量的取扱いのない総説型の論文を除外しました。方法としては、主に論文の方法のところに記載されているところに基づいて、研究実施施設、研究規模、研究対象職種、研究対象者、アウトカム、研究デザイン、統計学的手法について集計をしました。

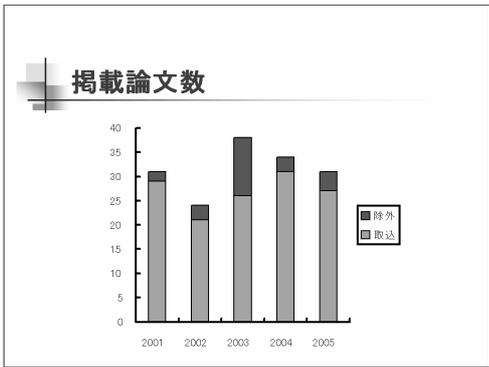
**対象・方法**

取込基準 2001年2月から2005年10月まで「医学教育」誌に原著・報告として掲載された論文

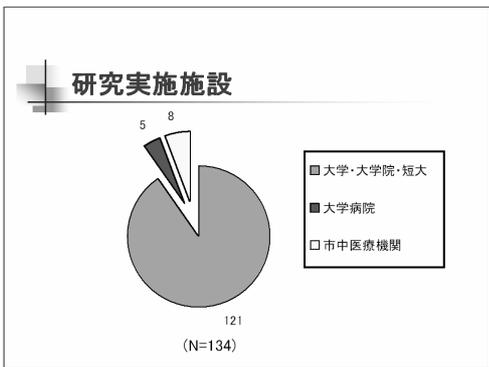
除外基準 データの数量的取扱いのない総説型の論文

方法 研究実施施設、研究規模、研究対象職種、研究対象者、アウトカム、研究デザイン、統計学的手法について集計

**14、掲載論文数:** これが2001年から2005年までの医学教育誌に載せられた論文掲載数のチャートなんですけども、全部で158ありまして、そのうち取り込み対象となったものが134、除外対象が24でした。2005年はまだこれから第6号が出ますのでやや少ない。



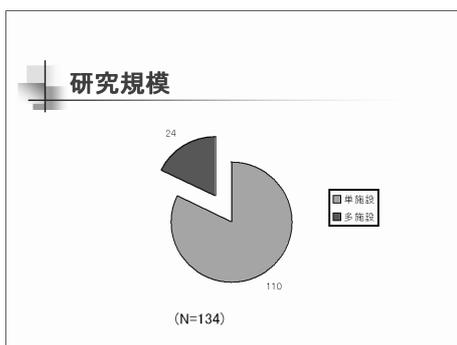
**15、研究実施施設:** 研究実施施設と書きましたが、研究実施者の所属施設ということでございます。大学・大学院・短大を含めました高等教育機関がほとんどを占めています。一部が大学病院や市中医療



機関でした。

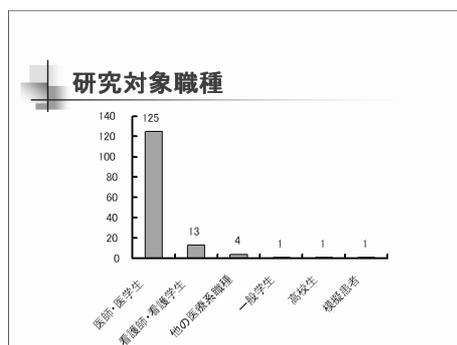
### 16、研究規模

研究規模としては1施設内では出されたものが8割ほど占めています、多施

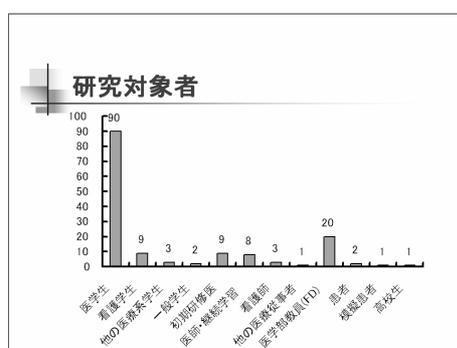


設間の共同研究が2割程度でありました。

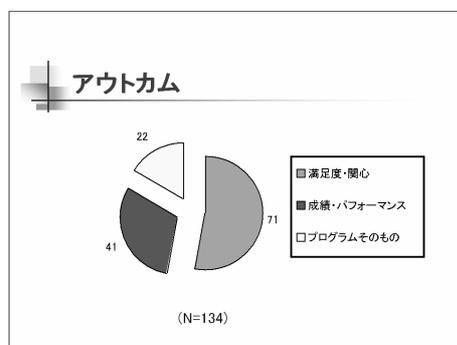
17、**研究対象職種**: 研究対象職種ですけれども、一つの研究で例えば医学生と看護学生の両方を対象にして調べたものがありますので、若干の重複がありますけど、やはり圧倒的に多いのは医師ないし医学生を対象としたものであります。そしてそれに続いて看護師・看護学生で、それから放射線技師ですとか医療系職種を対象としたものが続いています。



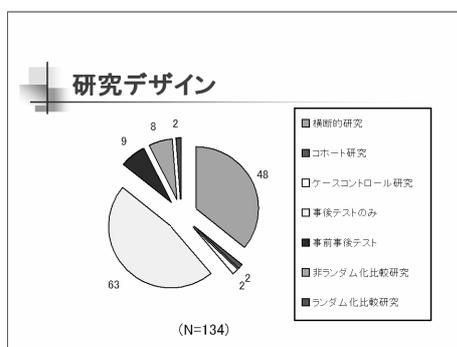
18、**研究対象者**: 続いて研究対象者ですが、これも対象者が多岐の領域にわたっているものもありまして、若干の重複はありますが、医学生を対象としたものが、看護学生、他の医療系学生左の4つまでが卒後教育を対象としたものでした。続いて初期研修、継続学習、看護職員、他の医療従事者、と続きます、FD 関連のスタッフが20でした。



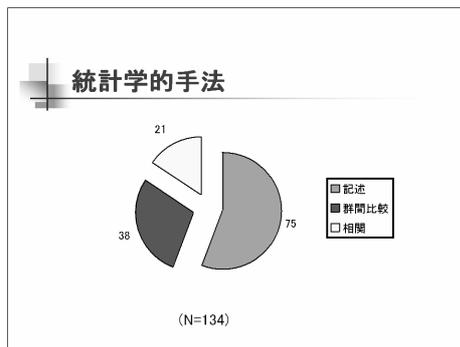
19、**アウトカム**: 測定対象のアウトカムですけれども、アンケート調査の形を行っているもののがかなり多くて満足度・関心度を調べたものが半数以上、対象者のパフォーマンスを調べたものが3割、プログラムそのものがうまく機能したかを調べてものが2割程度でした。



20、**研究デザイン**: 研究デザインですけれども上の3つを観察的研究としまして、したの4つを実験的研究と便宜的に分けました。これは研究者が対象に対して教育的介入をしたかしないかでわかっていますけど、観察的研究は3割強、横断的研究が非常に多い状況でした。実験的研究と、これは介入した後、事後テストのみのものが多くを占めていました、介入の前後で事前事後テストをしているのが9、非ランダム化の比較研究が8、ランダム化比較研究が全体の2でした。



**21、統計学的手法:**採用されている統計学的手法です。研究テーマに最も近い・・沿った部分で採用されている主要な統計学的手法ですが、記述によったものが75、群比較したものが38、相関が21でした。ここまでのまとめとしましては、2つありまして、1つは、研究の中では大学の医学部でされている研究が非常に多くを占めているんですが、医学生対象ですが、個人的には生涯学習ですとか・・そちらのほうもバランスが取れればい



いのかなと思います。2点目ですけれども、研究テーマをどのように定義するかによって、それにふさわしい手法や研究デザインが決まってくるので、例えば個々の観察データの研究から理論を作るのであればあるいはフォーカスグループインタビューのような手法を使って記述的な手法をとるのがふさわしいというふうになりますし、理論を個々のケースで確認するというものであれば実験的研究がふさわしいということになって、そういった研究テーマの種類によってふさわしい手法が決まってくるということだと思います。

**22、写真 23、写真:**私の発表はここまでなんですけれども、最後におまけですが、ダンディー大学をご紹介しますと、大学のメインキャンパスとそれから大学病院と2つが・・。丁度この2つの間に医学教育センターがあります。これが入り口で.....。クラスメートが7名います。・・最年長55歳・・平均年齢44歳・・〔足立先生終了〕



**24、教育研究手法そのものの吟味:** (大西先生)今のデータを見て、2枚だけスライドをまとめて作ったのですが、結局、今後どうしていったらいいのだろう、というところですね。今研究手法を吟味する場所、ってどこにあるのか、考えたのですが、1つには医学教育学会の年次大会で発表があります。今年もここで209でしたかね、あったのですが、主催校がふるい落とすようなことをやれば質をあげようという原理が

**教育研究手法そのものの吟味**

- 医学教育学会年次大会での発表
  - 主催校はまずふるい落とさないが...
- 「医学教育」誌の論文
  - 編集委員会は有効に機能しているか
- その他の医学教育研究発表内容の吟味
  - 一部は科学研究費報告書などの情報あり
  - 医学教育に関する科学研究費選別プロセス?

働くかもしれませんが、今のところそういうことはないですね。それから医学教育の雑誌の方、編集委員会は厳しくしたりはしているようなのですが、厳しいのが有効に機能しているのかどうか、正しい方向に厳しくしているかどうかですね、厳しくすることによって出そうという意欲が削がれるのであればマイナスですし、それからそれ以外の医学教育研究発表内容というのが科学研究費などの情報によりますと少しあるかもしれません。それからそういう科学研究費の選別プロセスというものです、いろんな科学研究費に医学教育分野のプロポーザルがでていると思うのですが、誰がどのような形で吟味しているのか、というのは窺い知れない領域ですけども。

**25、今後の課題:** 今後の展望ということで、医学教育研究を推進するためにどのような要因があるのかということ、おそらく、この2つくらいかなと、まず人的資源やその予算とかかなり上のレベルのことなのでちょっとここではディスカッションしても仕方がないことですので、国や大学で改善していただきたいことは多いのですが、上層部でのインセンティブは今ひとつだが十分働いていない、只、今医学教育研究がいろいろ出てきているのでポトムアップ的な改善はされるとは思います。そういったときにおそらく様々な研究技法の共有という

### ■ 今後の展望

- 医学教育研究を推進するために必要な要因の吟味
  - 人的資源やその予算
    - 国や大学のレベルで対応が必要
    - 上層部でのインセンティブは今ひとつだが...
  - 医学教育研究の様々な技法の共有
    - ワークショップはその活動の一つ
    - しばらくは草の根的な広がりを目指す

のが求められていくだろう、このワークショップはその活動の1つではあるのですが、でもおそらく上の方が十二分に動いていないので、しばらくは草の根的な広がりを期待するという、仲間集めですねそういった意味で皆さんも熱いディスカッションを繰り広げていただいたことは非常に心強いことです。ありがとうございました。

## 全体討論

Q:研究デザインのところで、現在医学教育研究にどのような展開、動向があるのでしょうか。

A[足立先生]:個人的な意見でよければ、データの数を増やすには多施設研究がよいのですが、データを測定する測定者が違ってしまう、どうしてもうまくデザインしないと、測定者間の誤差によって、大きく違ってくる。特にその教育関連では、必ずしも同じ構成概念でもって測定が行われなようなことも起こりえます。テーマのやり方によって変わってくる話なので、必ずしも質的研究より量的研究がよいというわけでもないのですが。

A(大西先生):私が思うには、他の施設に持っていっても非常に応用がしやすいような深みのあるテーマであれば、研究の意義が大きいと感じます。全国レベルでデータを取って研究をやった方がよいテーマもあれば、ある一部の学生に対して非常にコントロールされた構成された形で心理学的な分析デザインの研究をやった方がよい場合もあるわけです。どっちがいいのかというとケースバイケースでしょう。ただ全国レベルでデータを集めようとした場合におそらく、今、病院長会議が2年に1度行われていますが、カリキュラムのいろんなデータ集めていますけどそれが一番全国規模としてはまとまりのあるデータとなっているのかなと思います。そこに何かをくっつけてできるのであればなにかいいなあと思います。

Q:足立先生のダンディー大学は、授業料はいくらでしょうか

A:授業料はですね、マスターのコースで、8400ポンド、すなわち160万円ちょっとです。秋学期が9月から始まるのですが、秋学期と春学期とその後の研究の指導、全部含めてこれだけの値段、秋学期だけのコースですと3000ポンドちょっとですね。3400ポンド。秋学期・春学期含めてのディパタンスコースですと5100ポンド、100万円ほどです。大西先生のイリノイ大学はいくらだったんですか？2年くらいで卒業するとすれば単位をフルに取ったとして半期100万ちょっと、単位のない学期が10万円くらいですね。なかなか一概には言えませんが、1年目は学費免除になりましたから、そのときに単位を全て取ってしまえば安く済みます。

Q:足立先生のテーマは何でしょう

A:秋学期終わりのところで研究テーマを決めますので・・・これからということですね。

Q:先生のテーマは？

A:・・・で、あとは企業秘密です。

C:[大西先生]:日本で今教育研究を系統的に学ぶというものはないと思いますけど、そういった意味では海外のいろんなプログラムというのは魅力ありますね。

C:やはり、医学教育学の本格的な大学院コースがはやく日本にもできなければいけないという気がします。韓国ではもうすぐできるそうです。

Q:当初メールでいただいた内容には、生物医学研究との違いというのがありましたが、今の講演のなかでもそういったところをですね、説明して頂いたのですけれども、特に私はそういった生物医学研究しかしていなかったものですね。こういった教育の研究というのは馴染まないと言いますか、とつつきにののですが、今までのお話以外でも具体的にもう少しこういうふうにご注意をされるとか、こういったところに焦点をあわせるだとかこういうトピックスに焦点当てた方が、例えばその先程紹介されたような欧米のジャーナルとかに載りやすいとかですね、そういうような・具体的なことというのは何かあるでしょうか。

A:[大西先生]教育そのものがですね、政治や社会・文化的な流れですとかそういったものを受けて変わってきていますよね。例えばだんだん患者中心の医療をするようになってきて、学習者中心の教育をしないといけないというような意識が強まっていますよね、そういったやはり流れに段々着いていくと今、欧米で一番トピックスになっているのはプロフェッショナリズムですけど、プロフェッショナリズムに関する教育の研究というのは増えてきています。そういうとこについて行くということは一番重要なことではないかと思います。ですからあまり古いテーマを掘り返してやる、これ調べたいと、私なんかは診断推論に随分興味をもって、そのことをやろうと思って一番最初にイリノイ大学に行ったのですが、「いやあ、僕はこのことを30年研究してきました」という研究者とそのとき初めて出会ったのです。自分が思いつくようなことというのは殆ど大体されていたという印象があります。そういうところにとつくとあんまり得はしないですね。なかなか自分で成果が出せないというところがあります。

Q:学生の成績を解析したりというときがあると思いますが、入試関係ですね、そのデータをですね論文にしようとする大学の中のことなので、なかなか外に出せないかもしれない。そういう仕事に関してはどう考えればよいでしょうか。

A:[大西先生]やはりより普遍的に他の大学にも役に立つというふうなことが学内的にも納得がえるようなデータであれば、ぜひ出すべきですし、それは上層部の人も認めてもらえるべき、そういうレベルにいたっているかどうかの問題だろうと思います。それよりも、倫理的な問題、学内のデータで、学生にきちっとインフォームドコンセントをとったのかどうか、そういったところの問題、今個人情報のこともありますし、非常にナーヴァスになっておいた方がいいと思います。

Q:今大西先生がおっしゃったんですけども、Medical Education 誌の編集委員会では倫理的な問題ですね、ちゃんと倫理委員会通したかどうかですね。そのあたりを非常に強く問うようになりました。日本は採択率が高くて大体30%くらいだけど投稿数は少ないのが問題です。

## 第2日目 900-945

### 講演:教育研究を推進するために—教育心理学的方法をどう生かすか

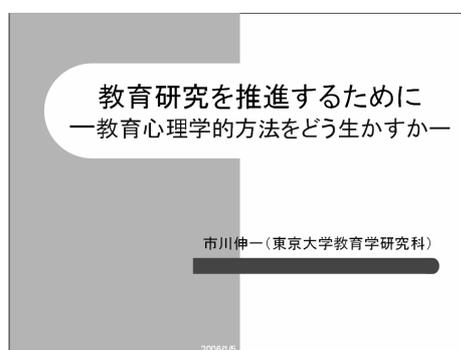
東京大学大学院教育学研究科教育心理学コース教授 市川伸一先生

市川先生のご紹介

今日は朝から市川伸一先生においでいただきました。現在、東京大学大学院教育学研究科教育心理学コースの教授でいらっしゃいます。プロフィール等はお配りした資料に書かれていますと思いますが、簡単にご紹介いたします。1977年に東京大学文学部を卒業されて、80年まで東大の大学院で心理学を専攻され、埼玉大学、東京工業大学を経て、東大に戻って教育学研究科の教授として活躍されています。主な著書としていくつか持ってきましたが、こういったものがございます。所属学会は日本教育心理学会を中心に活躍しておられます。では、市川先生、よろしくお願ひします。

今、ご紹介に預かりました市川と申します。文系の中でも心理学は理系に近いだろうということで、今日のお話を聞いていただきましたら、これは相当理系に近いことをやっているんだと思われるかもしれません。勿論心理学といってもいろいろですが、私が進学しました文学部の心理学専修課程、東大の中にも教育学部の教育心理学、文学部の中には心理学と社会心理学、この

心理学というのは伝統的な実験心理学をやるところで…。人間行動学というものがありまして、4つあるわけです。その中で最初は基礎的なことをかなりしっかりやりたい、やっていることも実験とか調査とか統計解析とかそういう理系に割りと近いようなところということで文学部の心理学を希望致しました。十数年間は基礎的な認知心理学をやっていました。心理学とは人間の知覚とか記憶とか推論とか問題解決のような知的な側面、知・情・意で言うなら、「知」にあたる側面の仕組みや働きを研究するという、そういう分野であります。次第に教育に関心が向くようになっていきました。もともと教育には関心があつたんですけど、最初は基礎的なことをしっかり、方法論としても身につけたいと思っていました。17、8年前になるんでしょうか、私にとって最初教育へのコンピュータの応用ということと個別学習相談を通じて一体子どもたちがどんなことが分からなくて躓いているのか、普段はどんな学習の仕方をしているのか、どうすれば分かるようになるのか、自分の教え方が通用することもあるし、しないこともあります。そのようなカウンセリングの学習版というようなことを心理学の分野でやってみたいと思い始めました。私にとって、コンピュータの教育法ということと個別学習相談を通じての認知研究というのが教育に関わる研究としてはじめにやっていたことです。その個別学習相談の研究は今でも続いていて研究会で臨床心理学の人たちがケース検討会、ケースカンファレンス医学の方たちもやっている、ケースカンファレンスをやるように自分が学習を通じていろんな子どもたちを見て面談をして個別指導をして、結果どうなったかを皆で検討しあう。どんな教え方をすればよかったのかなどを学校の先生も交えて話し合う、そういう研究



会が今でも続いています。そういうことをやっているうちに段々学校にも行きやすくなっていました。最初は教育心理学といっても学校というのは非常に敷居が高いところで我々が行っても大抵何の役にも立たない、ことが多いですね。すごく気がひけていました。ところが次第に個別相談なんかやっていると、こちらからの学習者が見えてくる面があります。1対1で見えていますので学校の先生が授業とテストだけでは見られないような側面がいろいろみられる。ある程度そういう知識をもっていきますとけっこうこちらが関われることもある、ということで段々学校に私たちも行きやすい場になっていきました。学校の先生も研究会に来てくれるし、私たちも研究会の学生や研究者も学校に行けるようになって来ました。そういうつながりが段々できるようになっていきました。そういうことをやりながら今教育心理学のコースにいるわけですがけれども、この教育研究を推進するために教育心理学的方法をどう生かすか、というお話をさせていただきます。

教育心理学とは一体何をするのか、ということをお聞きの方にもいらっしゃるかもしれませんが、こんなことをやるんだということをまとめました。教育心理学では4本柱というふうに言っています。戦後教職をとるのに大学では教育心理と教育原理、この2つを学ばなければならないということになりました。これはアメリカからの影響です。今はその名称は若干変わっていますが、教育心理学というのはそのときに非常に

#### 教育心理学とは何を研究するのか

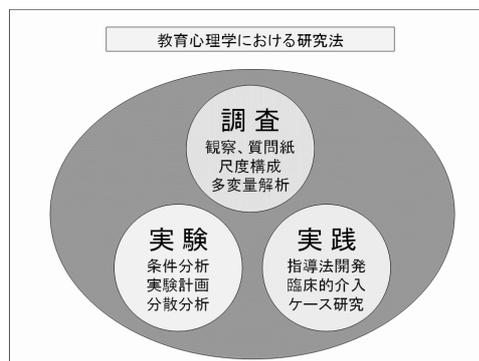
- 発達  
認知、情緒、社会性、性格
- 学習  
学習過程、動機づけ、学習指導
- 適応  
障害、不適応、非行、生徒指導、教育相談
- 評価  
心理測定、評価法、テスト開発、統計手法

2006/1/5

盛んになったわけです。どこの大学も教職科目として教育心理学をかならずやるわけですから、専任の教官も置くようになった。ということです。その時に教育心理学としては「発達」「学習」「適応」「評価」この4つのことを先生になる人に教えてくださいということになりました。発達の中には認知発達もあれば、情緒発達もあります、社会性の発達とはパーソナリティの発達もあります。学習というのは学習の過程、あるいは学習のメカニズムというようなことを研究します。あと、動機付け、学習意欲につながることで、それから学習指導、授業の仕方、それがどんな効果があるのかということも含めています。適応ですが、学校には障がい児といわれる子どもたちもいますのでその障がい、それから学校での不適応、それから非行とか、生徒指導、教育相談ということを取扱います。今、スクールカウンセラーといわれる人たちも入っていますが、学校の先生もそういう役割を担うこともありますので、カウンセリング臨床心理学に近いような分野です。最後、評価ですが、心理測定、評価法、テストを開発する、こともあります。もともと知能テストを開発したり、最近ですと学力テスト、あるいは学習に伴ういろんな意識を測定したりすることがあります。統計手法と書きましたが、教育心理学の中では統計手法の開発をするという人たちがかなりいます。大学入試センターではずいぶんそういう人々をかかえています、いろいろな大学にもそういう専門家がいます。特に東大では、そういう教育統計の専門家が2名います。2人いるというのは、これは大学のなかでは強いほうだと思います。教育心理学で開発された手法が他の分野、心理学以外の分野でも使われるということも結構あります。今日お話しする例えば因子分析というのはずいぶんいろんな領域で使われていますが、もともとは知能テストの分析ということで開発された手法ということになります。

す。

教育心理学における研究法ということで、3つ大きなカテゴリーに分けました。一つは調査、一つは実験、もう一つは実践です。先ほどちょっと紹介していただきましたこの『教育心理学法入門』というこの本は東京大学出版から出ていますが、この本は私も編者の一人ですけれども、こういう枠組で解説をしようというふうに編者のなかでも決めました。それほど教育心理学のなかでも、違和感のある纏め方ではないと思



います。わりとオーソドックス、このようなまとめかたがしてある本がどのくらいあるかは別として、この3つがありますよということです。調査というのは、できるだけ対象をありのままに捉えたい、現実場面での対象をありのままに捉えたいとして、観察したい、質問紙でアンケートのようなことも行います。それから質問紙をしたりやアンケートをしたりすると尺度構成を行う、尺度構成をするというのは、これは心理学の中にあると当たり前のようなのですが、要するにアンケートをやったらいくつかの項目をまとめて1つの物差しとする。その物差しの信頼性を高めるために複数の項目、同じようなことをきいていることもあります。信頼性を高めるために同じようなことを聞いてそれを合計することによって1つの尺度にする。というようなことを行っています。これは社会学等と比べると違ってきます。結構社会学はよく単項目でやります。何千何万のデータを取ってくれば、単項目で例えば社会教育の・・・がこうゆうふうに現れます、こういうことも結構説得力あります。ところが一人一人の例えば子どもを見たいというようなときに、1つの項目でこう答えたからこの子はこんな子なんだとしてしまうのは非常に危険です。むしろ、いろんなことをいろんな聞き方で聞いてその合計点として確かにこの子はこういう特性があると、いうように判断する。そのためにこの尺度構成というのを厳密に行います。それから調査でテストでもいろんな項目いろんな変数があります。その間の関係を調べるためには多変量解析というのを行います。回帰分析、それから因子分析、最近ですと共分散構造分析、というのが使われる手法です。実験の場合は、これは日常場面というよりは、日常にはないような状況をつくって、その中である変数が効果をもつかどうかを丁寧に調べる。他の条件は統制した上である条件を組織的に変化させるということをやっています。この条件が変わるとどんな影響が与えられるか、という条件分析があります。それから、かなりがっちりとした実験計画を組む、ということもよくあります。例えば教え方をABCと変えることによってどんな効果があるのか、教え方と先生の特性、として、例えば先生のパーソナリティとして3種類あれば、教え方×先生の特性ということで、9つの分類を作ってやってみる。そのときに使われる解析方法は大体分散分析ということになります。教え方の特殊効果があるのか、先生の特性の特殊効果があるのか、交互作用があるのかなど、分散分析が主流です。実践の場合ですが、例えば授業であれば実際に授業をしているなかでいろんな指導法を開発する、新しい教材を開発する、といったことがあります。個別のケースに対しては臨床的な介入、というのがありましたけど何か不適応を起こしてしまった子どもにカウンセリングをする、あるいは学習の不適応の場合でしたら、

私たちがやっているような個別学習相談もする、そのような介入という言葉、医学でも介入という言葉使われますか、(介入というのは治療的な意味合いですね)何かこちらから働きかけるということなので、介入といふとなんかいかにもお節介をしているような変な言葉と思うのですが、教育でも教育的介入といふのはよく使われます。一般の人から見ると介入したといふと邪魔しているみたいですが、教育の中ではむしろで何らかの指導場面で何かを指導する、何らかの働きかけをすることをいいます。その場合にはケース研究ということになります。量的な手法をとることもありますが、例えばある子どもに対して質問紙を繰り返すとかテストの成績がどう変化するか、量的な研究もありますが、割と質的な研究、要するに言葉による記述ですね。観察による記述、どのように変容していったかというような言語的な記述を元に考察していく、というようなスタイルが多く見られます。

一般論だけでは少しイメージが湧きにくいかと思われれますので、動機づけの心理学ということをして、心理学ではこんなことをやっているのかということ、典型的な古いものからをちょっと挙げますけれど、どんなところで心理学者は工夫しようとしているのか、というようなところを見ていただこうかと思います。動機づけといふのは心理学に中でもいろんな分野で行われます。むしろ学習心理学や教育心理学とか他の分野を対比していただくといいかもしれません。要するに“やる気”なんです、経営心理学のほうでは、労働意欲つまりどうやって、働くということに対して、動機づけを高めてもらうか、ということが大きなテーマになります。例えば何のために働くのか、いわれた場合にすぐ答えとして出てくるのは経済的な動機なんです。要するに働くのはお金のためだ、なぜお金が欲しいかといえればそれは「衣食住」というものに直結するからですね。要するに働くことによって収入を得て、それによって快適な暮らしをする、一昔前であれば当然の答えとしてそれが返ってきた、と思います。要するに明日のパンを食べるために働く、1日15、6時間働いていた時代もあったわけですね。しかし、時代が進むにつれて今度、親和動機といふのが強調されるようになります。これが1920年代30年代です。人間がなぜ働くかといった場合にむしろ職場での人間関係を重視するという考え方です。ある程度生活が豊かになってきますと、何も明日のパンを食べるために働くというばかりではなくなってくるわけで、むしろ職場で仲のいい同僚とかあるいは上司がいて、そこで何かをやりたい、

動機づけの心理学を例に

経営心理学から  
経済的動機：衣食住に直結する給与  
親和動機：職場での人間関係（ホーソン実験）  
達成動機：仕事を通じてのやりがい

学習心理学から  
外発的動機づけ：賞と罰／動因低減説  
内発的動機づけ：知的好奇心、向上心、理解欲求

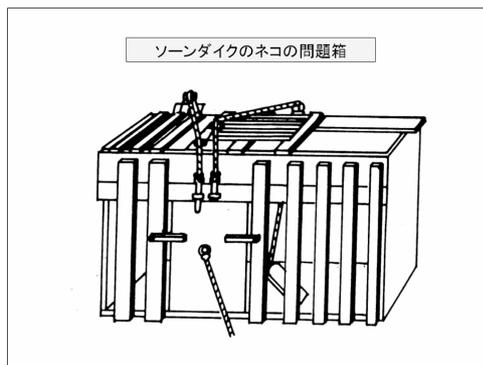
階層性理論  
低次の欲求が満たされると、より高次の欲求へ  
自己保存欲求、関係欲求、自己実現欲求

2009/1/5

アメリカのウエスタンエレクトリクス社というところにホーソンという場所に工場がありました。その工場ですべてどのような要因が生産高を上げるのに役立つか、という研究をしていたわけです。そのときに最初は物理的な条件に着目していた。例えば職場の照明はどれくらいが適切か、休憩時間はどれくらいの割合でとればいいのか物理的な条件に注目していたのですが、非常に物理的な条件が劣悪でもすごく生産高が高いということがある。そのグループを調べて見ますと、やはり同僚との人間関係と、上司との人間関係ですね、すると職場でいかにやる気をもって働くかといった時には人間関係的な要因がすごく大きいのではな

いかといわれるようになって、これが“親和動機”といわれるようになりました。もっと時代が進んできますと、今度は“達成動機”ということが強調されるようになってきます。つまり仕事を通じてのやりがいですね、何のために働くのか、仕事を通じて自分が是非やりたいと思っていることを実現したい。こういう欲求を職場に持ち込んでくるようになります。経営心理学の場合にはこのようにかなり現場に即したところでいろんな議論ができて、そういうことを高めるための職場が・・・ということがなされるようになりました。

それに対して、学習心理学というのも教育心理学の一部です。どちらかというともっと実験室的なところから研究は始まっています。外発的動機づけといきなりでてきたのですが、これは19世紀の終わりごろ、ソーンドイクという人が出てきたんですが、この人は教育心理学の父とか母とか、最初の教育心理学者とっていいと思います。この人はいくつかの仕事をやっているのですが、一つは基礎の研究でネコ



の問題箱というものを作って、学習研究を動物実験から初めたのです。これはですね、この中に猫が入るのです。最初は中でじたばたしています。ところがたまたまある仕掛け、この場合ですとあのペダルみたいなものを押すとドアが開いて、外にあるミルクが飲める、というふうなものです。こういうところにネコを入れると、最初はでてくるまでの時間が非常にかかります。

これは横軸が施行、縦軸が出てくるまでの時間です。すると最初は非常に時間がかかるのですが、試行錯誤しているうちに段々すぐに出てこれるようになる。終わりのほうになりますと入れた途端にペダルを押して出てくるような学習をするようになる、これは19世紀の話なのですが心理学が哲学から分化して、一つの科学的な学問になろうとしたときにどうすればいいのか、この行動心理学の考え方ではまず、学習

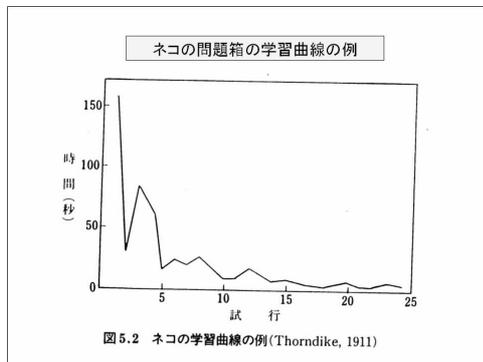


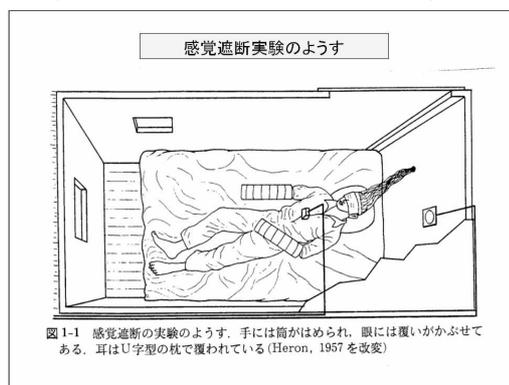
図5.2 ネコの学習曲線の例(Thorndike, 1911)

というのを何か知識を得たというような日常的な意味で使っているのではこれは科学になれない、要するに行動的な指標、動物にでも適用されるような行動的な指標の変化をもって学習というのを捉えるべきだ。この場合ですと、学習したというのは要するに出てくるまでの時間が短くなった。つまり学習というのは、心理学でも今でもかなり生きている定義ですが、何らかの経験によって比較的永続的な行動変容が起きたということ。要するに行動変容として、学習を定義する。そういう指標をどうするのか、さらにどういう条件であれば、外から観察可能な条件ですね。学習はスムーズに起きるか。ソーンドイクが非常に重視したのは、報酬の役割ですね。これは外に出てミルクがもらえる、というこれは動物にとって非常に大きな報酬です。こういうことがないといつまでたってもなかなか外に出てこないこともあります。つまり学習するためには、報酬が必要ということと、試行錯誤による練習が必要ということを示した。というか、当たり前と言えば当たり前

かもしれませんが、一つの大きな原理となった。それはこのころ進化論というのがかなり浸透して、人間にも動物にも共通の一つの学習原理があるといわれたわけです。

しかし、その後、本当に報酬というのがないと人間は学習しないのか、逆に報酬がなぜ動物にとっていいものか、といいますとそれはもし、今でもえさをもらうとか、そういうことがなかったら、どんどん身体が弱ってしまう、これは医学・生物学の方から借りてきた homeostasis という考え方なのですが、要するに、動物というのはただそのままの状態にしておくで段々むしろ安定状態を失ってしまう、血糖値が下がってしまうとか、何かそれまでの安定状態を保てなくなってしまう。保てなくなるとそれを保とうとして何かの行動を起こす、その行動がたまたまうまくいけば、報酬がもらえるということによって元の安定状態を取り戻す。そのために学習するんだ。これは一般的な原理とい

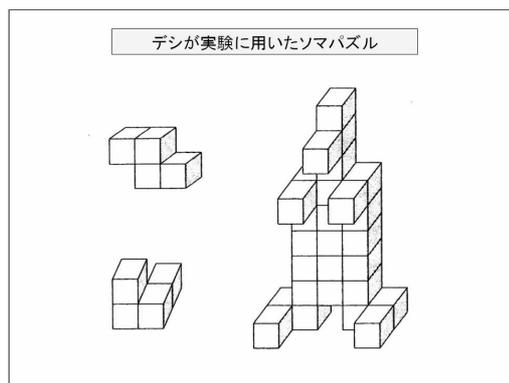
われたのですが、その後は、しかしそれが本当だろうか、という話しが心理学で随分出てきます。例えば人間がそれならば、外的な刺激とか与えないで例えば食欲とか、喉が渴いたら物が飲めるとかです、あるいは眠りたかったら眠ってもいい、生理的な欲求を完全に満足にさせてあげれば、そんなに快適かというこれはかなり極端な実験ですが、感覚遮断実験のようす、外からの刺激を与えない、内



的な生理的欲求は満足したい、そういう状況を作って人間を置いておくとうどうなるかという実験で。これはカナダやアメリカで1950年代、60年代に行われたのですが、実際にはこういう状態にして置くと、最初は高額なアルバイトなので喜んでやっています、応募者もけっこうあるのです。そのうち退屈さに耐えられなくなってきます、一日もするとだんだん退屈に耐えられなくなって自分で口笛を吹いたり、自分で刺激を起こして刺激を求めるようになる。2日くらいたつともうたまらなくなって出してくれ、お金はいいから出してくれというようになります。それでも頑張る人は一週間くらい頑張ってしまうようで、段々こちらが言う例えば幽霊の話とかでも信じてしまったりとか、簡単な計算もできなくなったりとか、頭も弱ってきます、後遺症も出るという話が出て、その後こういう実験は禁止になったのですが、極端な動因、先ほどの homeostasis のような動因が湧いてきたら、何か不安定になったら何か報酬をもらって安定状態を取り戻す。もし、生理的な安定状態が与えられていれば人間は学習などしないのだ、というモデルから、極端な実験として試してみて、どうもそういうものではなさそうだ、やはり常に情報を処理している、刺激が与えられて、情報を処理しているというようなことをやらないとむしろ人間は普通の状態さえ保てないのだ、という考え方が出てきます。それは内発的な動機づけというのですが、もともと人間は知的好奇心や向上心をもっている、むしろそういうものを適切に満たしてやる環境が教育の中でも必要ではないか。変に外的な報酬を与えると、かえってその報酬目当てに学習するようになってしまって、内発的な興味が削がれるということが言われるようになります。

これはデシという人が行った実験です。こういうソマパズルというのをを使って、大学生にやってもらいます。大学生もこのゲームを面白がってやるということは既にわかっています。このデシ自身

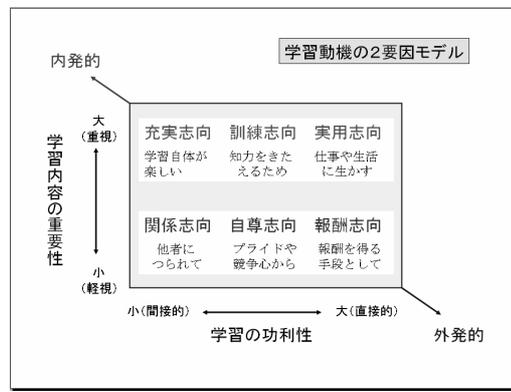
もすごく好きだったみたいですが、ところが、それまでの伝統的な外発的動機づけの考え方ですと、人間は報酬をもらってやることは喜んでやるんだ、という考え方でした。デシの場合はむしろ報酬を与えることによって、せっかくの面白さが、消失してしまうのではないかという仮説に基づいてこの実験を行いました。大学生を2つの群に分けて、一方のほうにまずこのゲームをやってもらいます。一つの群に



はそのときに報酬を与えます。一個作る毎に、こういうちっちゃなブロックを組み合わせで目標となる図形を作る立体パズルですが、一個解けるごとに1 \$出すという群と、報酬は全く与えられないという群に分けられます。その後に自由時間を設けます。自由時間というのはこれをやってもいいですし、雑誌なども置いてあって、雑誌を読んでいてもいい、とにかく何をしてもいい、内発的動機づけの強さというこの測り難いものを、その自由時間にどれくらいこのパズルをやるか、という時間で測定したわけです。つまり、自由にやっていると、その間の報酬はもらえません。ですから、やってもいいし、やらなくてもいい。しかし内発的な興味が持続していればこれをやるでしょうし、別に内発的な興味がなければ、ほかの事をしてしまうだろう。すると結果は明らかに最初に報酬などもらわない人たちはこれをやり続けた、しかし、報酬をもらった人たちは報酬がない自由時間にはとたんに止めてしまった、という結果が現れたのです。ですからもともと内発的な興味をもっているものに対して、へんに報酬を伴わせると、かえって外発的興味、知的好奇心等を削いでしまうのではないかと、これは教育界にもかなり大きな影響を与えたようです。一昔前ですと子どもは、いいことしたら誉めましょう、悪いことしたら罰しましょう、これは当たり前と言われていたのですが、あんまり賞罰で外から言わないほうがいい、これが極端になりますと、せっかくい成績とっても先生も誉めてくれない、親も誉めてくれない、本当にそれでいいのかということになるのですが、教育界というところは極端なところがあって、そういうのが浸透すると先生は絶対誉めたり叱ったりしてはいけません、親も誉めたり叱ったりしないようにしましょう。ということがすっといきわたってしまったり、表彰するなんてとんでもない、教育心理学会も論文賞というのをどうしようかやめようかという話がありました。半分冗談ですが…。しかしかなりそういう心理学での実験結果が教育界にも影響を与える、とりわけアメリカではあるようです。

最近はまだ少し柔軟な考え方になっていて、はじめは外発から入っても次第に内発に代わっていくんだというメカニズムも研究されています。この場合も教育心理学者が工夫しているのは、内発的な興味とか意欲というようなことを時間というような自由時間でこのような活動をするみたいな操作的定義に置き換えて測定しよう、というあたりかと思います。これは私の研究なんですけれども、心理学の中では普通よく外発、内発とよくいわれています。しかし、どうも内発外発だけでは子どもたちの興味を捉えきれないのではないかと、ということであまりそういう心理学での枠組に囚われなくて、ひとまず子どもがどんな動機をもっているのかをボトムアップ的にまず調べてから理論を作った、というものです。これは私が東工大にいたときですが、大学に入りたての1年生に「あな

たは高校までに一体何のために勉強をやってきたかと思いませんか」を自由に書いてもらいます。あるいは「人は何のために勉強していると思うか」ということも自由に書いてもらいます。そういう自由記述をまとめあげていって理論にする、私は最初むしろ動機づけの研究者ではなかったのですが、先ほどの学習相談のようなことをやっている、やはり子どもたちの動機というものに目を向けざるを得なくなります。特に学校の先生とか、すごく内発的な動機づけが高いと思うんです。高いからこそ教科に就いている。例えば数学の先生は数学が美しいあるいは数学で問題を解くのが楽しいむしろ内発的興味がある。大学の理学部、数学の先生もおそらくそうだと思いますね。あんまり何のためにということは意識ない、しかしそういう内発的な関心から授業を組み立てたときにどうも空回りしてしまう、ことがあります。生徒の方は一体それが何のためになるのかということがわからないとどうも興味が湧いてこない子も沢山います。例えば物理学とか工学とかにとっては数学というのは一種の手段ですね、数学それ自体が楽しいからやっているというよりは数学とは一つのツールとしてあって、どうしても実用的な興味を持っています。私たちにとっての英語、なんていうのも一つのツールなんですよ、学会発表をしたり、海外の研究者とコミュニケーションするためのツール、英文学の人は英語それ自体を美しさとか面白さでやっていると思いますが、でもそれだけで授業をやっていくとどうも空回りしてしまう、学習相談でも同じです。一体どんな動機をもっているかということを経験はそういう質的な分析、しだいにまとめあげていって、こんなようなものを作りました。6つに分類して、これを二次元に整理したものです。6つといいますと充実志向、学習自体が楽しいから、訓練志向、実用志向というのは仕事や生活に生かす、上の3つというのは割りとは教育界でもいわれています。下の3つこれも実際は子どもからよく出てきます。関係思考というのはみんながやるからなんとなく、とか先生が好きだから、他者との人間関係につられて、自尊思考というのは、競争心やプライドから、報酬思考というのは何か報酬を得る手段として、例えば成績がいいと小遣いを増やしてもらえとか、現役で大学にうければバイクを買ってもらえとか、そういうものになっています。横軸は学習の功利性と呼んでいます。右にあるものは割と学習をやれば得をして、やらないと損をする。要するに功利的捉えている動機が充実志向や実用志向です。縦軸は学習内容自体をどれくらい重要だと思っているか、上の3つはこの内容だからこそやる気がでる、内容にこだわったもの、面白いからこそやっているんだ、とか、この内容は役に立つからこそやっている、下の3つはあまり内容それ自体を重視していない。そういうふうに分類して左斜め軸が伝統的な内発外発というような分類で、そこにどれくらい内発的かどれくらい外発的かというのがでる。こういうのを最初はボトムアップ的に自由記述を纏め上げて作ったのですが、一端枠組を作ると、ここからは心理屋は尺度化するのが好きですが、それぞれの思考を表すような項目、というのを作っていきます。この場合でしたら、一つの志向について6項目を作って、どういう項目がいいか、本当にその6項目はその志向を測るのに妥当かと



いう分析をしていきます。項目分析というやり方でやっていきます。あるいは、因子分析やったときに、こういう志向が出てくるのがどうかというふうなことでチェックしながら項目を段々作り変えていく。この理論的な枠組が妥当かどうかというふうなことをやるわけです。動機づけということ为例にして心理学の中で実験とかこの場合でしたら調査になりますが、どんなやり方で行われているか…、実験とか調査に対して、実践というのがあります。実験とか調査というのはあくまでも、研究者が対象者を第三者的に捉えて、できるだけ客観的に研究しようというアプローチだと思っていただければいいと思います。実践というのは自らも関わる、一人称、二人称的な関わり方をすることが含まれています。実践と言ってもこれは秋田さんという人の分類をちょっと書き換えたものでこの本の中に、秋田さんが書いていますが、実践研究といっても広い意味狭い意味がありますが、これは相当広い意味の中でこういう5つのものがあります。非関与型の観察、要するにあるフィールドに入っていき、一時的なストレンジャーとして入っていく、よく授業に我々入るときに、よく透明人間になれと言われる。要するに邪魔にならないように、普段の授業の姿を捉えたいわけですから、自分が入ったことによって変わってしまわないように授業にそっと入る、影響をできるだけ与えないように入る、参与観察の場合には、自分もそのメンバーとして、その中に入る。文化人類学に見られますが、一メンバーとしてその中で生活をする。しかしある程度継続性があるもの。但しメンバーそのものというよりはストレンジャー、メンバーとして入るとはいうんですけど、異文化者、よそ者という形にどうしてもなります。もう少し関わるものとして、コンサルテーション、実践の間接的な支援者、例えば学校の先生に対して何か相談事を受ける、コンサルタントのような形ではいる。少し自分にも責任が出てくる。それから個別介入というのは特定問題場面での介入者、例えばカウンセラー、特殊訓練とか、指導ということを行う。私たちが勉強が不適応になった子どもに対して何らかの個別指導をするというようなこともここに入ります。日常的な実践ということで、これは学校の先生や親に普段から自分が何らかの教育をするという立場で関わっている、日常的に関わっている、全面的な実践者として、こういう人たちが自分で自分の実践を研究している、というスタイルになります。実践研究というのは、もともと教育心理学の中でも非常に研究としては認められにくい、ということがありました。どちらかといえば、実験あるいは調査による科学的な研究というのが模範とされる、その中でですね、教育心理学ではどういことがいわれてきたかという、教育心理学というのは不毛な学問である、ということの中からも外からもいわれ続けている。教育心理学の不毛性論文、不毛性シンポジウム、教育心理学というのはなんて不毛なのだろう、教育実践の役にも立たない、実践そのものも研究しているわけでもない、どちらかという実験室で実験をしたり、アンケート用紙やって分析したり、何の役にも立たない、そういう状況が戦後かなり続いていました。このシンポジウムも2年に一度くらいは学会の中で開かれていました。いくら開かれてもやはり改善されない、不毛性シンポジウムは不毛だ、という話も出るようになってしまった。それが10年くらいの前までの状況でした。ところが最近は不毛性シンポジウムが全く開かれなくなりました。なぜかという不毛だと言われぬようにずいぶん変わってきたからだと思います。心理学者がずいぶん学校にも入り込むようになりましたし、私たちも大学によって、子どもたちの個別相談をすとか、夏休みになりますと、文京区、台東区の子どもたちを呼んで、学習ゼミナール

のようなものを開く、そこで大学院生や私が授業もやって、データをとって、ということもやるようになりました。かなり教育実践の中に教育心理学者が浸透していくようになった。教育心理学の中身もずいぶん変わってきました。ひところはとも学校では使わないような学習材料を使って、これを覚えてください、ということをやっていたんですが、80年代からはまさに学校で使っているような例えば数学とか物理学であるとか英語であるとかそのようなものを素材とした認知研究が増えてきました。それはまさにどこで学習者がつまづいているのか、こんな誤解をしているというデータとか、得られますので、それは学校の先生にも役に立つ、ということでずいぶん役に立つようになんて変わってきた。

しかし研究としてそれを論文に出すと通るかというそれは結構難しい状況が続きました。それは、3、4年前から教育心理学研究という、こういう雑誌があります。この雑誌の中に実践研究というジャンルを設けよう、ということに私は関わっていました。15年前その話を学会でしたときに一掃されました。そんなジャンルは要らないと。ちゃんといいものがくれば審査して出してあげるからいいじゃない、といわれました。ところが載ることがどれくらいあるかというと10年に1本くらいしか載らなかった。その中でやはりどうしても審査員というのは、普通の科学的な研究を見る眼で、評価してしまう。それと実践研究をそういう基準で評価したのでは、なかなか載らないのではないかと、実践研究のそれなりの評価基準があっているのではないかと、そしてそれをもとにして、実践研究というジャンルにもっと論文を載せていくようにしようじゃないかと、実際学校の先生も随分最近学会に入るようになりまして、先生方も投稿したいわけです。ところがこれまでの教育心理学の基準でいうと通らない、やはりもうちょっと違う基準を設けよう、実践研究というのはというものだという定義はあります。定義はあるんですが、それでも評価はまちまちです。

私のほうで1つ実験というか調査というか、こういうことをやってみました。評価者によって、一体どれくらい実践研究に対する評価が違うものか、違うとすればそれは何が違うから違うのか、ここでは17編の公刊済論文があります。公刊済論文というのは、教育心理学研究に過去に出たものもあれば、他の雑誌に出ているものもあります。むしろ他の雑誌で割りと載せてくれる、例えば教育工学会ですね、そういうところでは実践的な論文が出たりします。あるいは各大学の紀要とか、そういうところはあまり審査を厳しくしませんので、そういう論文を17編集めてそれを教育心理学会の編集委員50人くらいいますが、その中で20名くらいが引き受けてくれました。この17編の論文を評価してもらって、実践研究として、評価したらどうですかを5段階、つまり採択から不採択、再審査、修正・・・の5

教育心理学における実践研究 (秋田, 2001)

- 非関与型観察調査
  - 一時的ストレンジャー、透明人間
- 参与観察
  - 継続的ストレンジャー、異文化者
- コンサルテーション
  - 実践の間接的支援者、コンサルタント
- 個別介入
  - 特定問題場面での介入者、特殊訓練・指導者
- 日常実践
  - 日常的・継続的な全面的実践者 (教師、親など)

2008/1/5

実践研究に対する評価 (市川, 1999)

- 17編の公刊済論文を20人の編集委員が評価
  - 実践研究としての評価と、通常の原著論文としての評価
  - 全体としては両者の相関は極めて大
  - 多くの評定者は、分化した評価をしていない
- 2つの意味での個人差
  - 全体的な甘さ・辛さ：個人ごとの平均
  - 評定の観点：評定者間の相関
- 評定者を変数とした因子分析
  - 分析志向の因子：実践過程、学習者の変化など
  - 開発志向の因子：新たな教育方法の提案、実施、評価

2008/1/5

段階ですね。もし、これが通常の原著論文だったらどうですか、ということで評価してもらいました。その2つの評価を17編の論文に対して、求めたのです。全体としてみますと、実践研究としての評価の行程と通常の原著論文の評価の行程はかなり相関の高い評価者が多かった。ということはいずれ、分化した評価をしてきていない、実践研究ならではの実践研究ではこういうのもありじゃないかと言ってくれる人が割りと少ない。個人差というのは、2つの意味での個人差がありました。全体的に甘い辛いというのがあります。それはその人の中での平均点を見れば分かります。しかし、それだけではなくて評定の観点が違う、これは評定者間の差をとってみれば分かりますが、観点がずいぶん違う、評定者間で違う、実践研究の場合には、評価者同士の相関をとるとすべてのペアについて行うのですが、無相関とか、ペアによってはマイナスというのもマイナス0.5というのとかもある、こっちの先生はいいというのがこっちの先生には悪いとかですね、ではなぜそのような違いがでてくるのか、ということで、因子分析を試してみました。20人の編集委員を変数、論文の方が評価される対象ということにして、割りときれいな因子が2次元出てきました。一つは分析志向の人がいますが、どれくらい実践のプロセスを細かく丁寧に記述しているか分析しているか、あるいは学習者の変化、そういう緻密な分析をしているものを高く評価する、という人たちこれが分析志向の高い評定者となりました。するとかなり直行する形で開発志向の因子がでてきました。これはどれくらいオリジナルで新たな教育方法を提案しているか、それを実施して、さらにそれが有効かどうかの評価までをやっている。こういう新しい何かを試み、であればそれを高く評価する、というのが開発志向の評定者です、これのどっちに重きを置くか、ということです。ずいぶん評定者間の個人差が出てきました。中には両方バランスよく考えてくれている人もいます。

その中の言葉として面白いのがあったのですが、この評定者 T さんというのは割りと中立的な方です。コメントとして、あなたはどんな実践研究が望ましいと思いますか、自由記述でいっぱい書いてもらいました。そのなかでこの T 先生がおっしゃることは実践研究というのはこういう研究が望ましい、つまり授業者とか実践者とか開発者自身の研究である。自分自身がやっているということですね。誰か他の

#### 実践研究とは一評定者 T のコメント

- ・授業者・実践者・開発者の研究であること
  - ・実証性・客観性ある、実践の評価をしていること
  - ・実践の重要性や、事例からの知見に普遍性があること
  - ・教材開発の場合は、有効性評価を含むこと
  - ・臨床事例については、教育とのつながりがあること
- 一般の原著論文は、  
客観的かつ厳密な評価・測定  
理論的背景の中で位置づけられ、理論的に貢献  
サンプルの普遍性、一般性

2006/1/5

人がやっているのを自分が観察しに行ったとかというのではなくて、自分自身がやっているというのが、実践研究であろう。実証性、客観性のある、実践の評価をしていること、これは自分の実践がいかにすばしかったかということだけを書いてあるものはあまり研究としてよくない、ということです、どういうところは有効であった、しかし、このような問題点もある。そこら辺はまだ限界がある、今後こうしていった方がよいであろう、というようなかなり客観的なコメントをして自分で記述していること、それから実践の重要性や事例からの知見に普遍性があること、確かに一つの授業とかカウンセリングでしたら一つのケースが論文になるわけですが、そこに何か普遍性がある、他の人にも役立つ内容である。教材を開発するという実践研究もあるのですが、それはどれだけ有効性だったかも含めて欲しい。臨床事例については教育心理学会でするので臨床心理学の学会と

いうのも別にありますので、教育心理学である以上教育とのつながりというのがあって欲しい、一般の原著論文というのはこの方が重視するのはむしろ、客観的かつ厳密な評価や測定、理論的背景の中に位置づけられる、理論的に貢献がある、サンプルの普遍性、一般性ということだけでも、実践研究の場合には、こういうことではないかといっています。あくまでも T さんの意見なんです、割と全体の中ではバランスのいいのではないかと思います。具体的にですね もう時間も私のしゃべる時間は過ぎてしまったのですが、お渡ししたいいくつかの論文があります。時代順に並べていただきたいのですが、「コンピュータによる心理学実験」、これは私にとっての最初の実践的な研究で、こういうものが教育心理学研究に出ています、最初は出してもらえとはあんまり思っていませんでした。駄目元くらいのつもりで出しました。私自身の授業、当時わたしは埼玉大学で経済学を専攻する学生に心理学の授業をしていました。普通の授業をやっている、どうも心理学というものにあんまり興味を持ってもらえなかったというか、おもしろがってくれない、むしろ実験心理学のおもしろさを分かってもらうために自分自身に実験の被験者をやらせてもらおう、そのときに実験者になるのはコンピュータです。コンピュータを実験装置として、実験者の役割をさせて、刺激を与えてそれから反応が起こる、結果を出す。自分は被験者になって、その時どんなふうにしてこの課題をやるかということを経験してもらう、それによって生徒も興味関心とか、データとしてはアンケート、それから出席状況、とか自由記述、そういう自分の授業について書いたものです。やはりこういう論文を出しますと、査読者は相当面食らうと思います。少なくとも自分の授業のことを論文にして投稿してきたというのはなかったと思います。やはり一人は全然だめだ、といいました。一人は資料ならいいだろう、ということで少し私もごねたのです。結局ある程度書き直して、資料に出してもらえました。2番目ですが、「コンピュータ・アートを利用した幼児教育の試行的実践と評価」、幼児教育、これはコンピュータグラフィックを使って、創作的な活動をやりながら・・・これは観察です。そのコンピュータ・アートで活動してる子どもたち、その場でも見ますし、ビデオもとりますが。それをいろんな評価基準というのが中に出ています、こういう行動のカテゴリー、基準を作って、一人ひとりの子どもをチェックしていく、どんなふうに変容が起きたか、というようなことをまとめたものです。これは観察法による行動指標として使われるものです。3番目は、「高校数学における授業観の構造と生徒・教師の対応関係」というもので、これは情報教育研究という雑誌です。尾城先生というのは私のところに1年間来ていた高校の数学の先生です。この尾城先生の問題意識は、いい授業とはどういう授業かということについて、先生の間でも随分意見が違う、生徒の間でも意見が違う、ですから数学教育の学会でもたぶん自分がいい実践だと思って発表をしてもしらせることが多い。あるいはいい授業だと思ってやっても生徒の方がしらせてしまう、そういうことから、授業観というものを調べてみたいということでやってもらったものです。質問紙調査なんです、最初に因子分析をやることを想定して、良い授業とはどんな授業か、授業ではどんなことをやって欲しいか、たくさん出してもらいました。先生からも出してもらって、生徒からも出してもらったものをまとめていく、ある程度整理してから、因子分析にかける。そうやって因子をまとめていく、学校によってどう違うか、先生と生徒はうまく対応しているのか、というようなことを質問紙調査によって調べた、という方法になります。4番目ですが、「高校生の英語学習

における学習動機と学習方略」というものがあります。これは先ほどの2要因モデル、動機づけの2要因モデルのもっと古いバージョンをここでは使っています。名前は少し違っていますが、ほとんど形は似ています。こういう学習動機が、どんな学習方略が起きるか、例えば英単語を覚える時にどんなやり方をとるかこれも因子分析でやる、3つのやり方がカテゴリー化されました。一つは、反復方略、ひたすら繰り返す、たくさん書いて、何回も単語帳をめくっています、ひたすら繰り返すという反復方略、それからスペルからイメージを思い浮かべる。スペルのイメージを思い浮かべて覚える。もう一つは体制化方略、これは認知心理学でも推奨されている方法ですが、要するに単語どうしの関連付けを行うパターン、私たちが例えば同意語・・・例文の中で覚える、例文のなかで他の単語と関連付けながら一緒にまとめて覚える、あるいは1つの単語が、どういう要素からなっているのか、一つの単語の中の構造ですね、例えば・・・。そういうかなり知識の体制化、構造化というものを重視した考え方。27ページというところを見ていただきますと、これは回帰分析です。単に回帰分析をつなげただけですが、もともと学習動機が3つあって、上の3つ、充実、訓練、実用この3つはお互いに結構相関がありました。内容関与的動機と1つにまとめる根拠になっています。下の関係志向、賞賛志向、報酬志向、今賞賛志向というのは、名前をちょっと変えています。内容分離的な動機、という内容関与的動機は、学習方略に影響を与える。学習方略と学業成績との関係はどうかというと、体制化方略をとるかどうかは英語の成績に結構影響を与えています。テストとしても3種類ありまして、イメージ化方略や反復方略というのはあまり寄与していない、やはり効果的で体制化方略をとるような学習者というのは、英語の成績もいい、というような結果を出しています。これは質問紙調査を分析することによる相関分析ということになります。最後の5番目の論文ですが、「概念、図式、手続きの言語的記述を促す学習指導」認知カウンセリング、私たちがやっている認知的な問題に関するカウンセリングということで、私たちがやっている学習相談は認知カウンセリングですが、その事例を通した提案と考察ですが、これは、かなり変わった論文なのです。実践研究というジャンルができて、最初は誰も投稿しませんでした。すごくニーズがあるといわれていたのでこのジャンルを作ったのですが、最初ですね、やっぱりお互いに敬遠していたと思います。一体どんな論文が出るのだから、どういうものを出せばいいのか、ということ、様子見しているような状態でした。で、私自身も出しました。教育心理学会というのは非常にある意味民主的な学会です。私当時理事長だったんです。最初はこの論文を出したときにまるで駄目という人がいました。とにかく匿名で審査していますから、全然わからないですけど。こんなのは論文ではないと、とてもこんな研究は載せられない、という人もいた。せっかく新しいジャンルを作ったんだから、むしろこれまでのジャンルではおおよそ採択されなくても何か載せる意味はあるのではないかというものを載せてもいいじゃないかと言ってくれる人もいた。結果的に何回か書き直して載ったのです。実践研究というジャンルを、その後をみますともう少しまともな論文が載っています。まともなというのはもう少しオーソドックス、むしろいわゆる科学論文でのフォーマット、論文の問題方法、で持っていくとき方法考察・・・これを踏襲したものが実践研究でも結構多いです。私はけっして実践研究の場合それに囚われる必要はないと思います。とくに臨床研修の場合は目的とか問題があって、クライアントを選んでいくわけではなくて、クライアントの方がくるわけです。ね、

授業の場合もこれを教えたいという目的があってやっているわけ、研究目的があって始めたわけではない。これを教えたいというのがあって、その中でいろんなことがおきて、あとから振り返ってみるとこれが研究としていいということでやっていくわけですから、必ずしもその形式に囚われる必要はないと思います。結果的にはこれが実践研究で最初の論文として出たですね。これをみたことによってこんな論文でも通るんだ、と多くの会員が思ってくれればいいと思いました。これでも通るんだったら自分でも出してみようかなと思ってくれればいいじゃないか、結果的にそれから実践研究が割りとコンスタントに出るようになりました。一つの号の中で、10編くらい論文がある中で1つか2つの実践研究がある。というかたちになりました。私のほうからひとまずお話をしたいと思った内容は以上なんですけれども、先生方はどんな関心でどんな問題をもっていらっしゃるか私の方でつかみきれないところがありましたので、ぜひいろいろ質問していただければと思います。私の方からこの会に寄贈させていただくという形でこの本(心理測定法への招待)とこの本(心理学研究法入門)の2冊だけはお渡ししましたので、皆さん何か参考に使っていただければ。「心理測定法への招待」の方がかなり入門的です。これから心理学を学ぶという人に基本的なことを知って欲しい。これ(心理学研究法入門)も入門のつもりで書いたのですが、結構難しいところもあります。私が読んでちょうどいいくらいの章もあります。ああ、なるほどそういう方法もあるんだと、結構新しい方法などが紹介されていて、卒論から大学院の修士論文を書くくらいまでの人に勧めています。じゃひとまずここで終わりにします。ご質問等よろしくお願い致します。

## 質問と全体討論

Q: とても面白い講演をありがとうございました。以前「学習と心理学」という本を昔読ませていただきました。先生は研究対象の児童や生徒だったり、多いようですが、今大学教育に関わっていらっしゃるの、そこで得られる経験と大学生、卒後教育の社会人ですね、そこでかなり共通する要素があったら教えてください。どういう風にご自身、感じていらっしゃるのか、教えてください。

A: もともと私は児童生徒に関わっていたのではないです。最初に認知心理学をやっていて、すごく自分でもコンプレックスとしてあったのは、自分には実践がない、ということだったんです。当時東大の医学部の先生と共同研究をする機会があって、お医者さんというのは、自分で臨床経験も持ちながら、研究もやっているという、そういう研究スタイルにあこがれみたいなものはあったわけです。ですからやっていることが足が地に付いています。心理学でそういうのを見てみると、臨床心理学は確かにそういうことをやっているわけですね。じゃ、自分でもできる実践はないだろうか、と考えたときに2つ、一つは大学での自分の授業を実践と考えるということ、それは先ほどの一番の実践研究に照らし合わせて考えてみて、自分でも実践やっているじゃないかと気づいたのです。自分が実践と捉えていなかっただけ。そこでいろんなデータをとったり、学生のレポートとか、出席状況をきちっと、成績つけるためじゃなくて、一体どうやると出席率が落ちたのか、これだってデータじゃないかと気づきました。一つは大学の学部教育をはじめたということと、もう一つは認知カウンセリングですね。認知カウンセリングの第1号というのは、やはり私の学生です。例えばコンピュータや統計学を教えていると、分からないという学生が自分の授業でいっぱい出てくる、それから

当時私は、SAS という統計解析の本を執筆してまして、やはり SAS が分からないという人が東大教育学部にもいっぱいいるんです、どこが難しいのかということを目録形式で聞いていく、たとえを使ったりしながら、こういう教え方をするとよくわかる、とういうようなことをまず大学生や大学院生にやってみた。同じようなことを子どもにもできないだろうか思って地域の子どもたちを呼んで学習相談を開いたということです。そういうふうにはいろいろな年代の子どもとか大学生大人に関わっていると、先生の今のご質問ですが、いろんなことがあってまとめにくいのですが、すごく大事なと思うことは、やはり相手の言い分を聞きながら教える、ということです。カウンセリングで、一方的にアドバイスだけしゃべっているカウンセラーっていませんよね、むしろ学習者に語ってもらう。それによって、こんな誤解をしていたんだ、というようなことが、その言葉の中に表れてきます。例えば中学生を教えているときに言葉の端々を聞いていると実は小学校で習ったことが相当誤解しているんだ、分かっていないんだ、ということがどんどん出てきます。自分の教えたことだと教えて分かったということで、子どもの答えは「うんなんとなく」、それでどんどん進行して行って、うまく教えた気になって、むしろ本人自身に語ってもらう中から誤解を発見して、それをそこに遡って、あんまり露骨に遡るとすごくプライドが傷つくので、やんわりと遡って、どういうふうには誤解を克服していくか、そのことというのは実は授業でも影響していたと思います。一般の学校の先生の授業でも、生徒が普段どれだけ分かったかということ把握するのは難しいのでついしていないと思います。一斉にわあっと説明して、次に問題を解かせるでテストしてみると「えっこんなことが分かっていなかった」と、そのとき、つい学習者に責任転嫁してしまって、ちゃんと授業やったからノート見れば分かるだろうって言ってしまったりします。それは自分はちゃんと教えているのだからできないのはあなたの責任だと言うことになります。確かにノートには書かれているのですが、生徒は難しいことをやっているのですから当然伝わらないこともあって、すると今でしたら今日わかったことわからないことをノートに書きましょう。これ小学校でも中学校でも行うようにしたのですが、ニコニコマークのところに、つまり分からないところには分からない印を付けるように、先生に対する質問カードを用意しましょう、すると、手を挙げて質問できない子でもそういうことは書いてきます。学習者の内的状態を捉える努力をした上で次の授業を組み立てる、これが多分個別指導でも授業でもいろんな形やろうと思えばできるだろう、大学生でも小学生でも共通するだろうと思います。違いは大学生の方はうまく表現してくる、心理学でいうメタ認知ですね、自分はどこが分かっていないか、ということ割りと言葉でうまく表現できる。小学生ですとそれがうまく表現できないので、こちらから把握しにくい。そこら辺の能力の違いがあります、むしろその能力の違いを育てて、最初分からないところにポストイットを貼るのもいい、として、小学校ではペタリンといって、分からないと思ったらペタリンを貼りましょう。教科書やノートにペタリンを貼って先生はそれを見ながら教えてあげる。というような、いかに学習者の状態を把握するかが重要でしょうね。

QC: 非常に面白いなと思ったのは、私たちがよく言っているのが、医学教育とか医療の世界はよく似たところがあって、授業と同じで患者中心、授業は学習者中心、今日先生がお話されたことはまさに医療面接だとかで丁度求められているようなことだったので非常に興味深いなと思いました。ありがとうございます。

Q:熊本大学の谷口です。先生のこの 2 要因モデルについて、聞き漏らしたかもしれませんが、関係志向ということについて教えていただきたいのですが、これは単に他者につられて、というかこれは自分が好きな人とか、親しい友達がいるとか、それにつられてという意味じゃなくて、他人との共同作業が楽しいとか、刺激を受けるというグループダイナミクス的なものを意味しているのでしょうか。

A:いくつかのものが込みになっているところはあります。最初先生がおっしゃったのは同調的なみんながやるからなんとなく同調して、というそういう面もあります。そういう項目もありますし、友達と何かやるのが楽しい。先生が好きだからあの先生に習う授業が楽しい、非常に感情的なもの、あまりにもその共同で学ぶことによって触発される、そういうものはこのなかにあんまり含まれていないです。結局、大学生とか大人だったらそれもあるが、もともと生徒から出たものを集計して、分類している限り、あまりそういうものが出てこなかった。ということもあります。ですから決して、これが全部というわけでもありません。生徒から出てきたものを整理してまとめた。ということなのでどうしても先生からみると、先生のおっしゃるその同調的なのは次元が低いように思われるかもしれませんが。けどやはりそういうものは無かった。実際に例えば中学生や高校生がそういう学習の仕方を学校で提言しているかというあんまりそういうことはないので、出てこなかった。という面もあると思います。でも共同学習によって、お互いに触発されあうのが楽しい、みたいなことで高校の授業ってそんなに動機づけがないんじゃないですかね、そういうものも出てこなかった。

Q:今日のことと関連してご質問したいのですが、この学習の二要因モデルというのはあくまでも教室という学校という枠組の中で、先生がつくられたものというものでしょうか(教室というか勉強ですね)ええ、私たちは医学関係者が多いと思いますけど、教室を段々と学習の場と考えなくなっています。一人で学ぶというようなことが学習である、そこで動機づけなど学習のプロセスを考える、その辺について教室を離れての教育をどう捉えればよいでしょうか。

A:これは教室に限ってはいないと思います。つまりあなたは高校のとき、ということですね。高校の時の勉強ということを考えると家で一人で勉強している、例えば受験勉強は一人でやりますよね、家で、というようなことがかなり、むしろ高校生が教科の勉強をしているというのが想定ですから、家で家庭学習をしていることもあるでしょうし、塾でやっていることもあるでしょうし、学校でやっていることもあるでしょう・・・

Q:小学校でいう総合学習みたいなものが医療の世界では主流となりつつあるわけですが、目的がその場合は、将来何に使うかわからない数学を学ぶような役に立っていないことを教えているわけではなくて、入学直後から、医療に関連したもう少し目の前のものを解決する動機づけを持たせた方がいいだろうと言われていますが、その辺は？

A:それは実用志向に入ると思います。例えば総合学習の場合、何かやりたいテーマ、自分が知りたいテーマがあって、そのテーマを追求するためには、むしろ基礎基本として学んでいかななくちゃ

いけない。というこういう流れで学習がおこることがある。私たちが研究していたら研究のプロセスの中で統計学の知識が必要になって学ぶとか、海外で発表したいのでそのためには、英語をしっかり勉強しないといけなくなったりとか。そのときには動機として、何か実用的な目的があって、そのためにこういうことを学んでいる。という動機です。それは実用志向の中に入る、これまでの学校ではわりとそういう面が少なかったと思います。小・中・高はですね。で、小中高の先生にしてみると、たとえば数学をどうやって生かすかと、数学って何の役に立つんですか、実は学校の数学の先生は非常に答えにくい、ですね。つまり、数学の先生自身は余り何かの役に立った経験がなくて、それ自体の面白さ、こういうと失礼なんですけど数学を何かの役に立つからと思って学んできたわけではないでしょうし、そうやって先生になってきた先生も面白さとか楽しさということを感じて数学の先生になった。その面白さを知って欲しいという方が多いんじゃないですか、その先生に「数学って何の役に立つんですか」と言われても困る。そんなことは考えなくていいんだよ、そんなこと考えなくていいくらい面白いんだ、だからむしろ経済学者にむしろ物理学者に、私たち心理学者にとってはあああの時習った行列の多変量解析はここで役に立つんですよと応用面はいえるわけですよ。実は9割方の方は応用している人なんですけれども、この1割の方が数学者になったり、学校の先生になったりしている。そういう面をもっと見せていくべきではないか、そのために今先生がおっしゃったような総合的学習とかですね自分がやりたいテーマがあって、それに付随して必要なことを学んでいく、そういう学び方を学校でも取り入れるべきじゃないか、それはモデルでいえば実用志向なんです。そういう側面は今までの学校では弱かった、国語もそうですね、国語学の先生に聞いている限り「国語やって何の役に立つんですか」というのはあまり答え返ってこない。だって源氏物語おもしろいでしょ、といわれちゃう。私たちに言われたりすると例えば学会発表するときにはしっかりとした発表する。とかですね、討論のとき受け答えをする。これはビジネスマンでも同様、どういふ国語力が必要であり何のために国語があるのか、という問いに答えることができる。逆に学校がそういう側面をもっと取り入れていくべき、というふうに思います。

Q:名古屋大学の西城です。学習動機の二要因モデル、例えばあまり学習に対してやる気の見られない学生に対して、時間内は自分からあまり積極的でなくていうとやるタイプ、夜はなるべく早く帰りたい、じゃなぜ入学したのかと訊くと「まあ、なんとなく」「親が医者だったから、ただ憧れた」。そういう子にはこのモデルと関連してどういふふうに対応したらいいかなと思ひまして。。

A:核心をつきすぎていて難しいですね。ただわたしはこのモデルは学習相談をする中からこういうことを考えようと思って、始めた研究だったのです。認知カウンセリングですね。そのときに質問紙としてできるいろいろな学校でもやってくれる中学校でも高校でもやってくれる一人ひとりに自分の結果を知りたい人はフィードバックするということもやっています。どの動機が正しいというのはないのです。どんな動機をもとうが本人の自由、勝手です。しかし、自分がどういふ傾向があるか、ということを見極めてこの説明もして、どうも自分の動機は片寄っているとかあるいは世の中にはこんな動機をもっている人もいるんだな、などを考えて、自分の動機を見つめ直すということはあってもいい、高校生くらいでしたら、ですから今のことで申し上げて一つには、少し自分の動機を内

省する機会がある、大学生だったらやれる、これは私の自分の授業でもそうです。自分の結果を知りたいという学生もいます。このなかで自分は実際に勉強不適應の子もっているわけですよ、あまりその子どもたちに最初から充実志向とか訓練志向とか実用志向とかですね、この3つは学校の先生にとって好ましいものですよ、普通は。学校の先生はまじめな先生が多くて、したの3つのような動機はけしからんと思いがちで、持つべからざる動機と思うんですけど…。勉強不適應の子どもに上の3つをひきつけてもやっぱりうまく行かない、むしろ入り口は下の3つでもいいんじゃないか、例えば幼少の子どもだったらシールを貼ったり、花丸を書いてあげたりとか、一種の社会的報酬になっているんでしょうけど、そういうことから入っていく、あるいは学校の勉強も嫌だといっている子に家庭教師がついたとき、まず関係志向、この先生とだったら勉強してもいいかなというふうに思って、もらう。そういうことを入り口にしながら、段々上の3つの動きが出てくるように考えればいいのですから、今の方のような場合でしたら、周りも家族も医者だし、自分が医学部に行ったという喜んでくれるたぶんに関係志向ということもあるでしょうし、医者になると儲かるといわれた報酬志向があるかもしれないですが、最初はそういうところから入っても、じゃまとにかく現場と一緒に試してみようよ、といったことでもいい先輩と一緒にいこうといってくれた、そこで経験することで段々医者としての学びが楽しくなるみたいなことが出てくる。それはいかに自分の成長にもつながっていくという感覚がもし持てれば、もっと上のような興味が出てくるような気がします。

Q:東京医大の大滝です。参考にさせていただける情報が多く、ありがとうございました。論文の形式のことで具体的にご紹介いただいたのですが、もう少し教えていただきたいのですが、1つは、教育心理学研究の投稿規程、実践研究のフォーム、フォーマットが一つと、それに関係していわゆる質的研究の論文にするときの形式なんですが、いろんなところで話題になっていると思うのですが、教育心理学ではどういうふうにお考えになっているか、お訊きしたいと思います。

A:まず実践研究というのはどういうものか、というのはですね。今の教育心理学会では定義のようなものを出しているんです。詳しくはこれをお読みください。どんなことを書いてあるかと申しますと、要するに実践研究というジャンルは授業研究とか学習相談とか、現実場面での実践を直接の対象とした教育心理学研究なんだ、ということですよということが書いてあります。要するに現実の場面ということは、あえて研究のために設えた場面ではなくて、現実的な場面ということですね。それから教育心理学的研究というのは、教育心理学の方法論とか、理論とかにやはり関連付けて欲しい、そういうことです。教育心理学的というのはすごく微妙な言い回しなんですけれども、何らかのデータを取ることと、発話記録のような質的データでもいいわけですよ。というデータにもとづいて分析するというか、何らかの教育心理的な方法論や理論に関連付けて欲しいと、ということですから、初等教育だけではないんだ、幼児教育、高等教育、社会教育でもいいと、実践場面で資料を収集して、実践の改善を直接目指すものが望ましい。ということですね。教育心理的な見地からの分析や考察にもとづく提言、実践の改善のための何らかの提言がされると、ということが書かれています。これを全部 AND で結び付けてしまうとかなり難しいと思うんですね。さっきも言いましたが開発志向が強いとどうしても分析が甘くなってしまう傾向にあります。分析がものすごく詳しいものはちよ

っと開発的視点から捉えたときにそんなにいいものはないというものがあって、全体を満たすのは殆ど難しいので、こういうことを考慮しながら書いてくださいということになります。実際にはフィールドワークのように自分が実践しているわけではないけれども、実践場面に来て詳しい記述をして書いた、観察したというものも含めて。そういった質的な研究が投稿されるととにかく、査読者は割れるんです。全体的に言っているのはまず、例えば膨大な観察の記述みたいなものが表れたとしてそれをどういう形で整理していったのかという手続きをある程度きちっとしてください、例えば KJ法みたいなものがあります。最近グラウンデッドセオリー法というのが結構あります。流行りつつあるんですけど、そういうどんな手続きでこういう膨大な資料をどう整理していったかという結論を最終的なモデルとか理論にいたったのかその手続きを明確に記述するようにしましょう。原則としては追試可能なことだと思うんですね、追跡調査、同じ人がやってもう一回やっても確かめることができるものが望ましい、そして記述して欲しいということ…、それから論文の形式として、ある査読者はですね、これ私たちがだした別の論文で、きちっと問題目的、それから方法、考察と、こう書きなさいと、研究である以上それができるはずだ、むこうが言ってきた、私はそれにこねたんですよ。つまりそれをおしつけなくてもいいじゃないか、実践するということに何も最初に問題目的がきちっと決まってからやっているわけではなくて、長い実践の中からやったものが変化していくこのことこそ実践の醍醐味というわけで、あまりそれに囚われることはないのではないか、これは編集委員会でも合意事項なのですが、ちょっとむこうも黙っちゃいまして、結局私たちが載せてもらいました。ただ、全体としてみると実践研究の半分以上はそのフォーマットをとっています、とれるものはその方法、形式で書いています。査読者のほうも絶対だめとはいわないけれども、そういう一定のフォーマットで書いてくれるほうが読者にとっても査読者にとっても読みやすいじゃないかと思えます。

Q:秋田大学で仕事をしています。今作業療法士を養成する立場にいます。作業療法士の学生たちもやはりなんとなく来てしまったとか、医学部に行けないからまあ作業療法、リハビリにきたという学生がいて、結局そういう学生たちの動機づけやそのための授業をどうしたらいいのだろうか悩んでやっているところです。まとめていくのが一つの実践研究だという…さきほどの方とだぶるところがあるのですが、先程、形式に囚われることは無いというのは私も同感なんですけれども、特に目的問題があって、ケースがくるわけではない、その点そのとおりだと思います。ところが実践研究の中で事例研究といった場合に…学生たちに教える場合、後戻り的なものではなくて、やはり最初意識的に問題をもった上で、事例をとっていくべきではないかなとした教え方をしているもので、そういう言い方をすると、なんかちょっと矛盾してくるようなところがあります。本音としては同感するのですが、実際にまとめて論文にしていくときに、目的問題があってケースがくるのではないということと、事例としてまとめていくにはどうしたらいいか。

A:私もそれが悩みですけども。ケースというのは、沢山持つわけですよ。それぞれのケースは確かにこっちの研究目的にあわせて来てくれるわけではないですよ。こっちの目的に合わないからお断りするわけにもいかないですね。とにかくむこうのニーズに応じていろんなケースはみて

いくのですけど、そういうケースを発表しようという段階になった時には、なんでそのケースを発表しようとするのか、というところでこちらにも研究発表の目的もしっかりある、ですね。これはこういう研究の文脈に載せてみると、このケースはすごく意味があるんだ、と、ですからイントロダクションにはそういうことを書くべきだと、書いていいと、こういう研究の文脈があると、そして今回のケースはこの研究の文脈で考えるとこういう意味をもっているんだ、実際にそのケースを書いたうえでこういう研究の発展にこういう形で寄与すると思われまます。というくらいでなんとなく折り合わないだろうかなという気はしています。

Q:九州大学の吉田と申します。かなり現実的な話になってしまうのですが、こうやって実践的研究ということで論文が書けるということが、大体なんとなく分かってきたのですが、まず、それを出すジャーナルというのは日本では先生のところの雑誌しか、(そんなことはないです!)教育工学会でも受け付けてくれる?

A:教育工学会はわりと実践性というか、どう役に立つかということ結構重要視する学会で、しかも教育工学という学問はもともと伝統があったわけではないですから、新しくできた分野というのは、あまりそういう型に縛られずに、人工知能システムを作りました、そういう話もでる一方、学校でこういうことをやりましたという話も出ます。むしろ多くの人を学会に入れておきたいというそういうことがありました。そのために教育工学は結構、懐が広い、受け付けてくれるところがあります。

Q:それで同じようなことがですね、海外のなかでもあるのかどうか、それはかなり、現実的な問題例えば学位審査をするときにその例えば医学部では欧文じゃないととてもたぶん教授会で承認が得られないんじゃないかというところがあるので、もう日本に出せばみんなの役に立つのにも思っても、...そういうところはどうなのですか。

A:やはり海外のメジャーな雑誌で、こういう学習に対するケース研究的なものを載せているのはみたことがないです。もしかしたらそれもよほど最初から拒否しているわけではない、昔の教育心理学会と同じで、と思います。教育心理学研究と同じで、いいものがくれば載せますよというスタンスだと思います。じゃいいものというのはどういう基準でいいものといっているかというところについて査読者は普通の科学的論文の要件というものを相当要求します。ケースじゃモノはみえないでしょ、もっとこういう研究を少なくとも10人くらいやってから出してください。同じようなケースがそんな都合よく、くるわけない。逆に臨床心理学ではケースレポートを載せることもあるんだと思いますけど。いわゆる...こういう研究では海外の研究でもケース研究的なものは多い。

Q:ということは戦略的に考えると、結局国内で欧文雑誌を作ってしまう・現実としてはもうそっちの方が速いのかもかもしれませんね。

A:そうですね。それは確かに戦略的にも欧文のジャーナルに出すほうが価値があるとみなされているわけですね。人にもよるのですが、ケース研究はもうむしろ日本人に読んでもらいたいし、こういうのは日本語で書くというのがいいんじゃないかと、医学ほど英語で書くことが求められるわけではないかもしれませんが、普通の科学的な論文は海外にだすと、世界的なユニバーサルなものとして欧米の雑誌に載せた方が評価が高い。自分でも研究した中で、この認知カウンセリング

のことを歴史とか書いて本に載せましたが、同じことを例えば書いても雑誌に出しても多分載らないのではないかなと思います。本だからこそ載せてくれた、非常に難しいですが、いずれにしてもそういうニーズはむこうでもきっと起きてくると思います。むこうでもそういう研究している人はいるはずなので・・・、そういう人たちがどんどんむこうのジャーナルにも論文をだすようにして、日本人も出していけるといいのではないかなと思います。

少し時間もオーバーしてしまいましたが、有意義なディスカッションでした。教育学あるいは教育心理学とそういったところに専門分野の先生方と我々医学教育、医療者教育の面々がこういったいろんな対話をするという経験は昔はあったみたいですけど、この世代ではあまり無かったことじゃないかなと思います。(市川先生には)非常に柔軟な考えがあって、もともとは非常に科学的な心理学的なところから実践的な応用的な面に入ってこられるという流れを教えてくださいました。今後研究のあり方を考えると、そういったことにまで色んな示唆がありました。今後も教えていただきたいと思っております。今日はどうもありがとうございました。

## グループ討論3: 仮想データに対する統計解析方法

(大西より)下の課題に対して、各グループで取り組んでください。エクセルや SPSS のデータを見て単に議論していただいても結構ですし、ある程度分析を加えて報告していただいても結構です。

課題: どのような統計学的解析を用いるか

2つの Dataset (PassFail, PSQ) があります。これらを解析するための統計学的解析方法を選択して下さい。

1) PSQ (佐賀大学小田康友先生によるデータ提供)・・・表2

44 人の学生が外来で患者との医療面接を経験しました。表1の患者満足度評価票 (PSQ) を用いて、以下の3つの点について確認したいと考えています。

- (ア) PSQ の各項目間の点数に差があるか
- (イ) 学生の性別による点数の差があるか (性別: 0-女性, 1-男性)
- (ウ) 2週目には1週目に比べて進歩がみられるか

表1. 米国内科学会による患者満足度評価

1. あなたへ丁寧にあいさつをしましたか。あなたへの呼びかけが適切で、親しみを感じるものでしたか。気難しい、または、失礼と感じる態度はなかったですか。
2. 態度が高慢だったり、子供を扱っているような感じだったりせず、対等な立場で接していましたか。
3. あなたが話をしている間、さえぎることなく自由に話をさせ、十分聴いてもらえましたか。また、質問は適当なものでしたか。
4. あなたに対しては、十分な関心をもった態度で接し、退屈な様子や、あなたの話を無視するようなことはなかったですか。
5. あなたに質問がないか尋ねましたか。質問したことについて、わかりやすく説明しましたか。あなたの質問を無視したり、説明がなかったりしませんでしたか。
6. あなたの症状や治療について話をするとき、専門用語をつかわず、わかりやすい、簡単な言葉を使っていましたか。

「不十分」「まあまあ」「よい」「非常によい」「特に優れている」の5段階で評価。

2) Pass Fail (東京大学大西よりデータ提供)・・・表3

マレーシアの医学部で 154 人の学生が3年生への進級試験を受け、進級できなかった学生が 30 名 (進級不可=0, 進級-1) 出ました。進級できなかった成績不振者の原因を入学試験内容に求め、性別、人種 (CHI: 中華系, MLY: マレー系, IND: インド系, OTH: その他)、面接 (2~3名の面接官による独立採点、面接を受けていない者も)、入学時成績 (1が最も良く、4が入学最低ライン)、入学時英語成績のいずれが最も成績不振者の要因となっているかを調べたいと考えています。

表2. PSQ データ

ID	年齢	性別	PSQ1	PSQ2	PSQ3	PSQ4	PSQ5	PSQ6	PSQ 計	週
25	26	1	4	4	4	4	4	4	24	1
25	26	1	3	3	4	3	4	3	20	1
25	26	1	3	3	3	3	3	3	18	1
25	26	1	4	4	4	4	3	3	22	1
25	26	1	3	3	3	2	3	3	17	1
25	26	1	4	4	4	4	4	4	24	2
25	26	1	3	3	3	3	3	3	18	2
25	26	1	3	4	4	3	3	3	20	2
25	26	1	3	3	4	4	3	4	21	2
25	26	1	4	4	4	4	3	4	23	2
25	26	1	3	3	3	3	3	3	18	2
43	26	0	3	3	2	2	3	3	16	1
43	26	0	5	5	5	5	5	5	30	1
					⋮					
142	24	0	2	2	3	3	3	3	16	1
142	24	0	4	3	4	4	4	4	23	2
142	24	0	4	3	3	4	4	4	22	2
144	24	0	4	3	4	3	3	3	20	1
144	24	0	3	3	3	3	3	3	18	1
144	24	0	4	4	4	4	4	4	24	1
144	24	0	4	4	4	4	3	2	21	1
144	24	0	3	2	2	2	2	2	13	1
144	24	0	4	4	4	4	4	4	24	2
144	24	0	3	2	3	3	2	3	16	2
144	24	0	3	3	3	3	4	3	19	2
145	24	0	5	5	4	4	4	5	27	1
145	24	0	5	5	5	5	4	5	29	1
145	24	0	4	4	4	4	4	3	23	2

表3. Pass Fail データ

ID	性別	人種	年齢	面接1	面接2	面接3	入学成績	英語語彙	英語流暢	英語文法	英作文	進級
15	M	CHI	19	37	32	34	1	8	4	4	24.5	1
16	F	CHI	19	33	34	34	1	8	3	3.5	18.5	1
17	M	CHI	19	28	30	29	1	6	3	3	25	1
18	M	CHI	19	37	36	37	1	8	4	4	20	1
						⋮						
255	F	MLY	19	29	31	27	3	7	4	4	20	0
256	F	MLY	19	30	32	37	2	7.5	3.5	3	23	1
258	F	MLY	19	41	32	34	4	8	4	4	24.5	1
262	F	MLY	19	21	21	24	3	6.5	3.5	3	21.5	1

### 全体発表と討論

**3班:** PSQを最初に解析しまして、そのあと Pass Fail の方に行きました。仮説がアイウありますのでそれに沿ってやってみたんですけど、全体の平均を出して、質問項目毎に差があるなど、質問5と6がやや低いな、ということについて見て行きました。その後男性と女性を比較すると、男<女

女性が良くて、男性が少しおとると、ずっと平均値をみて・・・統計で、全体の傾向を、把握していくという作業をやりました。それから1週目と2週目では、やはり2週目が改善が見られる(1週<2週)、次に1週目と2週目で男性女性の向上の差はどうなんだろうかというようなことをやっていく条件がどんどん群が増えて行って、こうやってひたすら群が増えていくのだろうかというところで、終わりました。

		PSQ1	PSQ2	PSQ3	PSQ4	PSQ5	PSQ6	PSQ 計
女性	mean	3.70	3.67	3.70	3.70	3.52	3.55	21.84
	SD	0.69	0.76	0.70	0.70	0.75	0.76	3.80
男性	mean	3.44	3.53	3.61	3.58	3.30	3.40	20.87
	SD	0.68	0.66	0.73	0.71	0.81	0.76	3.58
1 週	mean	3.54	3.52	3.54	3.51	3.33	3.40	20.83
	SD	0.71	0.71	0.73	0.71	0.76	0.75	3.69
2 週	mean	3.63	3.69	3.77	3.78	3.51	3.56	21.93
	SD	0.68	0.73	0.69	0.69	0.80	0.78	3.71

それで、使った検定方法に関しましてもこれは、尺度としては、順序尺度であり、間隔尺度ではない。単純に数値化していいのであろうかという問題がある、データとしてはあまり1や2に付けたくないという心理が働いてか、345に集中していますし、検定方法をどのように用いればよいかということで、検定に関する手法を議論を少ししましたけれど、そのあたりで時間が大分なくなりました。恥ずかしながら私はPSQのデータの提供者だったんですけども・・・これは学位論文だったのですが、この時は結局t検定をひたすら繰り返すということとはよくないということで、ANOVAを使いまして、PSQの項目ごと、そして性別そして1週、2週の3つのことを含んだ3WAY ANOVAを使いまして。それによると項目ごとに差があるとか、1週2週で改善はあるけれども、改善率で男女差がある、交互作用があるということが、傾向があるということが見えた経験的に話して、そちらに関する議論は終わりました。なんで先程申しました疑問、順序尺度を間隔尺度で用いること、それからこれをノンパラのデータをパラメトリックでやること等に関しての、それでいいのかということに関しましては、大西先生がいいといったということで、それが一番の根拠じゃないかと、国際的に通用するようになるにはあと10年ほどかかると思います、日本では大西先生がいいっていい・・・私には説明できませんでした。

	年齢	面接1	面接2	面接3	入学成績	英語語彙	英語流暢	英語文法	英作文
進級不可	19.73	33.24	32.76	32.20	2.87	7.22	3.47	3.35	22.42
進級	19.81	33.83	33.77	33.11	2.15	7.66	3.73	3.67	22.85

次に、このマレーシアの国際医大の合格不合格のデータを見ましたけれども、これも同じように記述統計を出していきまして、全体の傾向を概観していったというようなことをやりました。合格者と不合格者の人のデータを比較して行って、みていったのですが、まず気がついたことは、人種による差がすごく大きい、チャイニーズは通っているのですが、マレー系が不合格が圧倒的に多いとい

うことに着目しました。そのあたりで時間が尽きました。なぜ、マレー系が地元なのに落ちるのかということに関しては、母国語がマレー語であって、この大学は全部英語で、講義をしていて試験も当然英語で行われるという問題もあるのか、あるいは私立の地元の金持ちが入ってくるとか・・・成績は悪くてもマレー系がはいれるのかということ推測して解決には至りませんでした。

Q: データについての確認ですが、初診の患者さんで患者さんは皆違う人を研修医は皆同じ、..

A: 全く初診患者だけです。初診患者さんで、それで答えてもらっているのです。この評価表は学生にフィードバックは全くしていません。調査の許可はもらっていますが。研究協力を。見せるとこれに応じた面接をしはじめますので..。

#### どのような統計学的解析をもちいるか

- PSQ
- 患者さんは全て初診と想定する

Q: 1週目2週目とに差があるとは考えられませんでしたが、学生にとってですか

A: 2週間実習がありまして、連続してやっているのです。大体15人から20人の新患の外来患者を診ます。1週目に診た患者と2週目に診た患者です。分けたということです。それ以外に改善の度合いを見る指標がなかったので、そういう分け方をするしかなかったということです..。

#### PSQの各項目間の点数に差があるか

- 全体が一様かどうかの検定
- 差があるとすれば、どの様な差があるか
- 各項目の平均点を出す 比較をする
- クラスカル・ウォリス検定をして
- 10パターンのt検定

Q: 患者さんの年齢とかは？

A: 年齢とかありますけども、ここには載っていない、年齢とか性別とか職業の有るなしとか、教育歴とかそういうことも調べています、ただそれによってあまり差が出なかったですね。差が見られたのは職業の有無でした。職業のある人がやはり、評価が厳しいというぐらいで。

#### 学生の性別による点数の差があるか

- 対応のない2標本の代表値の差
- 個人の重みをつけて調整
- 個人の代表値を用いて合計点の差、項目の差を対応のない検定
- もしくはカイ二乗検定

**2班:** もうすでに答えを言われてしまって言い難いのですが、こちらのグループでは質問項目の妥当性をまず検討することで、逆転項目がないと、本来は因子分析をして、すべてのベクトルが 同じ方向にベクトルが向いているかを確認して、一部に..が入っていることを確認しました。5段階評価ということで、順序

#### 2週目には1週目に比べて進歩がみられるか

- 進歩をどう計るか
- 1週目と2週目は対応しているとみなす
- 一人の学生の各項目ごとの1週目の平均点と2週目の平均点
- 一人の学生の1週目の合計点と2週目の合計点
- 対応のあるt検定

尺度で今言われたようにこのグループでは、前提として、初診の患者で、全員すべての学生が違う患者を診たということ、という点でディスカッションをしました。最初のPSQの各項目間の点数に差があるかということで、各項目の平均点と分散をみて、各項目の有用性、項目で分布が、どのような分布をしているかというような全体が一様かどうかの検定、差があるとすればどのような差があるか、を出すということで、一応ここではノンパラメトリックで対応のない標本の代表値の検定ということで、マン・ホイットニーの検定を行うということをお話ししました。各項目で5項目の中から2項目ずつ選びますので $c_2$ になりまして、10パタンのt検定を行う。次の学生の性別による差があるかということで、これに関しては、学生の平均値を出して、年齢的な要素がある程度揃っているか対応を検定する必要があるかどうか・・・それで挙げたのが学生の平均年齢ですがこれは、対応のない2標本の代表値の差を検定して、例えば特定の女子学生がいい点をずっととり、男性がずっと悪い点をとると学生の性別差があるということに繋がるので、個人の、学生個人個人の平均点を出して、それを代表値として合計点の差を出す。それから、項目を対応のないt検定という形でマン・ホイットニーの U 検定を用いて、この人は  $X^2$  を4つに分けて  $X^2$  をやった・・・最後に2週目は1週目に比べての進歩があるかということで、進歩をどう測るかということですが一応1週目と2週目がランダムに患者が無作為にあたったとして初診患者が無作為にあたったとして、1週目にあたった患者と2週目にあたった患者はほぼ均一であるということで対応する標本としました。各1人の学生の各項目ごとの1週目の平均点と2週目の平均点と差の検定を行いました。対応がある2標本の・・・t検定を行う、・・・つきの・・・検定が・・・。

Q: かなりたくさんの方の、検定をやっていくとなると、一つ一つに関しては  $\alpha$  エラーは5%だったとしても全体としてエラーがどんどん増えていってしまうというふうな問題をどのように考えられたでしょうか？[大西先生]

A: 設定を5%を繰り返す場合は、さらに厳しくしなければいけないというふうなディスカッションが生まれて、その時は・・・ボンフェローニの調整をするとかですね、そんなことも必要だねと出ていました。

**1班:** うちの班はPCに嫌われてしまいPCエラーが生まれて、時間でPSQしかできませんでした。まず、考えたのが代表値の選択をどうするか、学生によりましては数名の患者さんから評価を頂いているのですがある学生は1名、・・・かなりばらつきがある。それでどうするかということですが、最頻値でみるか平均値でみるかもう一つ中央値でみるということもあるのですが、中央値でみるにはデータ数が厳しいかなと、で、今回は平均値を採用しました。それはなぜかと、こちらの勝手な思い込みもあるのですが、PSQが患者の満足度

- 代表値の選択
  - 扱う患者数が異なる
    - 最頻値 or 平均値      平均値を採用
  - 前処理
    - 評価項目の関連を因子分析する
    - 要因項目のグループ化ができるか
- ア      被験者内要因のある分散分析
- イ      対応のない t検定
- ウ      対応のある t検定

評価としてきちんと成立しているのだというふうに使っているとすれば、この各項目は基本的には正規分布しているのだろうと、母集団は正規分布しているんだろうで今回やっているわけですから、一応平均値を使っていいのではないか、ある学生は1人の患者さんしか評価されていない、他の学生は複数の患者さんからの場合には、複数の場合には学生のデータそのものは信頼性が高くなるのですが、片方は例えば10人位の患者さんから、片方の学生は1人の患者さんからだ、その差どうするか、10人の患者さんから得られたそれをランダムに勝手に1人だけとる、そうすることによって、データとしては、すべての学生がすべての患者さんからランダムにあてられたというふうにすることができるかなど、その次ですけれど、まずこれらの6項目がすべてある同じベクトルを持っているかどうかをみる必要があるだろう、それは因子分析をしまして、きちんと負荷量がある一つの因子に全部高い負荷量が集中しているかということを見る必要がある。それが、一応これも分析しないとわかりませんが、一因子なのだろうと考えましてその場合にはファイルにありました合計得点を使うことができるだろうということです。まずPSQの各項目に差があるかどうか、これは被験者内要因のある1要因の分散分析を行うということでもいいのではないかと、いうふうに考えました。次の性別による点数の差があるか、ということですが、これも対応のないt検定、あと1週目と2週目のデータを込みにして考えるか、それとも分けて考えるか、込みにする場合でしたら単純に今でているようなものでいいのですが、場合によっては1週目と2週目を要因にして、考えた方がいいだろう、そうするとある場合には被験者要因の2×2、2要因の2×2水準被験者内の分散分析になるのではないかと、2番目も同じ、アは2×6水準、イは2×2水準の2要因の被験者内要因の分散分析になるだろう、ウですが、2週目に1週目より進歩が見られるかということですが、これは1週目2週目ということを変数として対応のあるT検定で検定すればよろしいのではないかとということになりました。

c: PSQを使って何を評価するかということはやったときに非常に悩んだのですが、各学生のA君B君C君の能力を評価するにはバラつきがありすぎる、とっている患者の人数にも差が有りすぎて10人取っている人もあれば1人しかとっていない人もある。たまたま相性が悪かったのかもしれない。となると全く使えないとなるし、もともとPSQ自体が個人の評価に使うには20例くらいの患者さんに評価してもらわなければ一般化できないということ、がマニュアルにありましたので、結局何の評価に使ったかという、総合外来実習のプログラムの評価に使った、学生はどうでもいい、200何十人の患者さんから評価を受けるわけですがけれども、総合診療部の実習がプログラムが学生の能力を向上させるかどうか、という評価で1週目2週目の学生の進歩、..いうことでやったということだけ追加しておきます。万事うまくいく。

c: こういうスタイルですと介入の前後でとっているのが本当に介入に効果がありかどうかはわからないので、多分僕だったら、2週目の学生と1週目の学生を違う学生にして、t検定をやると思いますが。違う学生も入れる。実習を受けていない学生との比較をやらないと..比較群のないプレポストだけをやっているということですね..

## 講演:大西先生から「統計学的解析」について

今、実際のデータを見ていただいて、コンピュータの使用になかなか問題があったところもあるようでしたけど。どうもすみません。

**1、統計学的解析:**まず統計学的解析をやる上でこちらのグループはかなり系統的に記述統計をだして、傾向をある程度大まかな傾向をみておいて、そのあとで分析的な統計を取っていくといった流れを作っていました。おそらくそこが一番基本として大事なところなのではないか、まず大まかな雰囲気を見ておかないと間違った方向に変な方に非常に複雑な解析だけを行おうとすると問題があります。

**2、パラメトリックとノンパラ:**それから今話題に出いましたパラメトリックにするのかとノンパラメトリックにするのかという話題ですね。本来の数理的な考え方からいきますとパラメトリックな手法は正規分布、等分散を仮定して利用可能、ノンパラメトリックな手法は分布型を問わないという意味で、その方が安全性が高いということですね、結構いろんな海外の臨床試験なども使われています。でも検出力は劣る欠点があります。それから、いわゆるパワーを計算してサンプルサイズを事前に検討するようなことはあまり簡単にできないことがあります。複雑な解析をしようとするとそういうことが非常に難しくなって、結構心理測定というのはこういうノンパラのようなデータを扱っているのですが、ノンパラにされていないですね。非常にここは逆説的なところはあると思います。複雑な統計学的な手法を使わないといけないので、ノンパラだとうまいかないからやはりパラメトリックにしよう、そのような発想の転換と考えていい。そこの1つの理由付けとしてNが30を超える場合に、パラメトリックな手法でも頑健性があるって大丈夫なんだ、みたいな話で..頑健性からまず問題ない..。

**3、平均の差をみる統計解析:**これはどの本にも書いてあります。対応のないデータか、対応のあるデータかというこの見分けが大事だろうと思います。

**4、カテゴリ間、2変数間の相関:**今日あまりデータ関係ないですけど、相関解析というのがありますが、

### 統計学的解析

- 記述統計
  - 割合、分布(ヒストグラム)、平均、標準偏差...
- 分析的統計
  - 平均の比較
  - 相関やカテゴリ間の傾向
  - 回帰

### パラメトリックとノンパラ

- パラメトリックな手法は正規分布、等分散を仮定して利用可能
- ノンパラメトリックな手法は分布型を問わない。しかし、検出力には劣る。濫用は控えるべき
- nが30を超えるような場合、パラメトリックな手法でも頑健性からまず問題なし

### 平均の差をみる統計解析

検定目的	パラメトリック	ノンパラメトリック(順序尺度以上)
対応のない2標本の代表値の差	平均値の差のt検定	マン・ホイットニーのU検定
対応のある2標本の代表値の差	平均値の差のt検定	符号検定 符号付順位和検定
対応のないK標本の代表値の差	一元配置分散分析	クラスカル・ウォリス検定
対応のあるK標本の代表値の差	混合型分散分析	フリードマンの検定

### カテゴリ間、二変数間の相関

- 二区分変数、名義変数:  $\chi^2$ 検定
- 順序変数: スピアマンの  $\rho$  ( $r_s$ )
- 順序変数: 散布図、ピアソン積率相関係数( $r$ )

$\chi^2$  検定とかスピアマンとか二区分変数、とか、順序変数・・・ありますが、それぞれの変数によって使い分ける。このあたりは教科書で済みますね。2つ目のパスフェールに関しては回帰分析の考え方が入ってきます。

**5、回帰分析:** 多数の独立変数がある場合、1つでもいいですが、従属変数の予測式を作ることも基本的な考え方です。従属変数yが間隔変数なのか二区分変数なのかによって線形回帰分析かロジスティック回帰分析かの使い分けをする。これはカテゴリーで4つくらいある場合、非常に難しいのですがそれを間隔変数とするよりは最近、三区分、四区分の数値として特殊なロジスティック回帰分析をやる方法があるようです。パスフェールに関しては最終的な従属変数が・・・パスフェールの2つですからロジスティック回帰を使うということになるんだらうということになる。

### 回帰分析

- 一つ、または多数の独立変数によって、従属変数の予測式を作る
- $y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$
- 従属変数の種類と回帰分析の関係
  - 間隔変数: 線形回帰分析
  - 二区分変数: ロジスティック回帰分析

**6、入学選抜と予測妥当性:**

**7、分散分析=Analysis of Variance:** おそらくその生物医学的な研究では、あまり分散分析というのは出てこないですね、3群位あって、それぞれの比較をするというときに出てくるとおもいますが、2要因3要因同時にということはありません。先程も言いましたけど繰り返しのあるデータかどうか、要するに被検者内の要因・・・、今回もこれに従いましたけど、一応エクセルや SPSS のような計算式のソフトウェアでは繰り返しのないデータというのは縦に重ねるということですね、性別であれば0か1でずうっと縦にならんでいる。繰り返しのあるデータは横に並べるPSQの1~6というのは横に並んでいる、それは項目というのは一応 within subject それぞれの被検者の中でいろいろな要因がありますから、繰り返しのある・・・

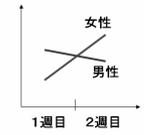
### 分散分析=Analysis of Variance

- 繰り返しの有無を間違えない
  - 繰り返しのないデータ (between subject)
  - 繰り返しのあるデータ (within subject)
- Excel, SPSSでは
  - 繰り返しのないデータは縦に重ねる
  - 繰り返しのあるデータは横に並べる

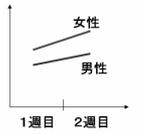
**8、2Way,3Way ANOVA:** 二元配置、三元配置といいますが、こういう複雑なデザインになってくると交互作用というのが出てきます、たとえば1週目と2週目でどう関係だった、こちらだったら1週目2週目に関係なく、性別に関係なく伸びていく、こちらは、男性が少し鈍いですが、一般的に伸びていく。これは交互作用なしということです。こちらの方は女性の方は伸びているけれど、男性はあまり伸びていない。そういう結果が出たときは、交互作用ありという言い方をします。このグラフが重なるとか、非常に大きく開いてしまるとかは交互作用あり、ということになるわけです。

### 2way, 3way ANOVA

- 二元配置、三元配置の場合、交互作用も確認



交互作用あり  
男性は伸びないが、  
女性は伸びが著しい



交互作用なし  
男性も女性も  
2週目には伸びた

**9、期末試験で不合格になった学生の入学成績と入学時の要因:** ちょっと簡単に後のほうのパス

フェールのもとになった、これは今年医学教育学会で発表した内容です。説明します。

**10、背景と目的:** セメスターの12345となっており、半年毎になっていて、2年半の臨床前のカリキュラムで再試験が1と3と5とあって、ここは留年とかもあります。今回155名中30名が留年したということで、入試方法の検討を考えた方がいいのではと、留年者の要因検索をせよということ学部長から突然依頼されて、やることになったのです。

**11、国際医学大の入試制度:** 入試制度ですね、ややこしいのですが、いろんな入学許可試験というのがオーストラリア入学許可試験、英国入学許可試験、いろいろあって、イギリスだと A-Level というもので、科学系:BBC 以上だったらいいと、AAA だったら非常によい、英語検定試験はこれもいろんな試験を受けていく、TOEFL だったら550点以上になっています。他にマレーシア英語検定試験でもいいわけです。それから面接を加味して半年毎に約180名のスロットがあって、試験を受けた人がこういう点数をとって、よければどんどん席が早い順に埋まっていくんです。これは西洋式のところだとそうです、東洋式のものみたいに同時に試験をして、一気に決めてしまうというのは東アジア特有なのではないか、入学時に英語能力試験を受験してもらいます。これは後々英語の補助的な教育が必要かを見分けるためです。入学願書にはいろんな性別とか年齢とかデータがあります。

**12、方法:** 方法として、155名を対象にして総括評価の可否を従属変数にしました。独立変数が年齢・性別、最初まずt検定と $\chi^2$ 検定でそれぞれの要因に関して差があるのかをみました。この辺はいろいろな方法論があるようですけど、 $p < 0.2$ である程度差がありそうだと思うものは要因を選択して多変量解析に取り込む、何でもかんでも入れてしまうと、あまりにごちゃごちゃしてしまうので、ある程度整理をする。最後はロジスティック回帰分析で $p < 0.05$ で有意になる要因を選択しました。

**13、結果:分析Part 1:** まず、1つ目をやりますと、このような感じになります。人種は非常に大きく関与して

## 背景と目的

- 国際医学大(マレーシア)の臨床前総括評価



- 再試験は各1回、Sem3, 5は留年あり、Sem1は留年なく強制退学
- 2004年3月Sem3の155名中30名留年
- 入試方法の検討目的で留年者の要因検索

## 国際医学大の入試制度

- 各種入学許可試験で一定以上の点数
  - 英国A-Level: 科学系BBC以上
- 英語検定試験で一定以上の点数
  - TOEFL(筆記): 550以上
- 面接を加味して半年毎に約180名の席を早い順に埋めていく
- 入学時に大学が提供する英語能力試験を受験
- 入学願書に各種人口動態的データ記載あり

## 方法

- 2002年後期入学生のうちSem3に進級した155名
- 従属変数: Sem3総括評価の可否
- 独立変数: 年齢(20歳以上or未満)、性別、人種、入学資格試験成績(試験毎に2段階)、学内英語テスト(流暢さ、文法、語彙、作文)、面接での点数
- 分析part1: t検定、 $\chi^2$ 検定で $p < 0.2$ の要因選択
- 分析part2: ロジスティック回帰分析で $p < 0.05$ で有意になる要因選択

## 結果:分析part 1

入学時要因	P値
英語: 流暢さ	0.006
英語: 文法	0.01
英語: 語彙	0.04
人種	0.001>
入学資格試験成績	0.007

いるだろう、入学資格試験というのも1~4ですね、いろんな試験があつてある程度範疇分けていますけど、ここでは関係ありそうです。それから・・・この時点で面接は省かれます。・・

**14、結果:分析 Part 2:** 入学試験の成績というのも1から4でいろんな試験があつて、ある程度わけていますが、英語が関係ある、ということです。ロジスティック回帰をやると英語の中で英文法、人種ではマレー系のところで・・という結果が出ました。結局この2つの要因というのが後々成績に絡んでるだろう。

**15、面接の意義?** : 面接の意義というのが、問われたわけです。2, 3名の教官が入って1名の学生を同室で約20分面談する方法、志望動機とかカリキュラムに関してどれくらいしてもらっているのか、対話能力はどうか、入試面接FDというのを一度やりましたが、限定的な効果しかなかった印象があります。面接の点数の信頼性というのは、非常に期待が薄い事がわかりました。1対1面接にしてステーション数を増やした方がいいというマックマスターのデータがあります。

**16、国際医学大でのアクション:** このようなデータを見て、結局国際医学大学では、PBL & 講義の Hybrid curriculumをやっているということで、英語教育を充実させることをやりました。それから、特定人種、マレー系ですね、が受ける頻度の高い資格試験、英語資格試験の最低基準を挙げて入学しにくくした、

**17、まとめ:** こうした対策を打つと言うときに、いろんなこういう統計学的なデータは有用だろうと思います。ありがとうございました、ご質問は・・・ないようですねアンケートをお願いします。また、お一人ずつ 20 秒間ずつ感想をお願い致します。

## 閉会

皆様から、非常に温かいフィードバックの言葉をいただきまして、ありがとうございます。我々委員会自身は本年度で12月に任期を終えるのですが、来年度も本委員会が存続されることを祈っております。その節にはまたこういうワークショップを設けたいと思いますので、また皆さんお越しいただければと思います。よろしくお願い致します。どうも今日はありがとうございました。

### 結果:分析part 2

入学時要因	P値	オッズ比95%信頼区間
英語:文法	0.04	1.03~5.15
人種	0.001>	(マレー系)4.32~31.5

★英語の各要因は互いに関連性高く、人種と入学試験の種類、点数も互いに関連性高い

### 面接の意義?

- 2, 3名の教官が同時に1名の学生と同室し、約20分面談
  - 志望動機、カリキュラムに関する知識、対話能力(英語)、応用力、個別の質問
  - 入試面接FDは行われたが、効果限定的
- 点数の信頼性は期待薄  
➢ 1対1面接でステーション数増やす方が良い?

### 国際医学大でのアクション

- PBL&講義のHybrid curriculumと同時に英語教育を充実させた
- 特定人種が受ける頻度の高い入学資格試験と英語試験の最低基準を上昇させた

### まとめ

- 入学選抜時の成績や人口動態的データを縦断調査に用いることで、選抜方法の改善に役立てることができる

## 第1回医学教育研究技法ワークショップ ポストアンケート

(2005. 11. 27)

1. 今回のワークショップの内容はあなたのニーズに合っていましたか？  
とても合っていた 5 まあまあ合っていた 18 どちらでもない 1 余り合っていなかった 1
2. 次の学習目標に到達するためにこのワークショップは役に立ちましたか？  
とても役に立った 4 まあまあ役に立った 18 どちらでもない 3 余り役に立たなかった 0
3. 以下の各プログラムは役に立ちましたか？
  - ① 概要説明  
とても役に立った 6 まあまあ役に立った 16 どちらでもない 2 余り役に立たなかった 0
  - ② シナリオに沿った研究計画  
とても役に立った 8 まあまあ役に立った 12 どちらでもない 5 余り役に立たなかった 1
  - ③ 講演: 研究立案, 実施の流れ—生物医学研究との違い  
とても役に立った 6 まあまあ役に立った 13 どちらでもない 6 余り役に立たなかった 0
  - ④ 懇親会  
とても役に立った 3 まあまあ役に立った 5 どちらでもない 4 余り役に立たなかった 1  
不参加 13
  - ⑤ 市川伸一先生講演「教育研究を推進するために—教育心理学的方法をどう生かすか」  
とても役に立った 17 まあまあ役に立った 8 どちらでもない 1 余り役に立たなかった 0
  - ⑥ 仮想データに対する統計解析方法  
とても役に立った 1 まあまあ役に立った 19 どちらでもない 4 余り役に立たなかった 2
4. 今回のワークショップについてお答えください
  - ① 時間配分は適切だったと思いますか  
そう思う 5 少し思う 14 どちらでもない 5 思わない 2
  - ② 進行は適切だったと思いますか  
そう思う 13 少し思う 9 どちらでもない 3 思わない 1
  - ③ 今回のプロダクトを自施設での医学教育研究に取り入れ(続け)ようと思いますか  
そう思う 13 少し思う 9 どちらでもない 1 思わない 1
  - ④ 参加費は適切だと思いますか  
そう思う 15 少し思う 6 どちらでもない 5 思わない 0
  - ⑤ 今回のワークショップを他の人に勧めたいと思いますか  
そう思う 12 少し思う 11 どちらでもない 2 思わない 0
  - ⑥ 今回のワークショップの続きがあれば参加したいと思いますか  
そう思う 20 少し思う 5 どちらでもない 0 思わない 0
5. 今回のワークショップで、よかったと思われる点を書いてください。
  - 統計解析方法が十分理解出来た
  - シナリオに沿って検討できた点, 教育心理学の話
  - 様々なバックグラウンドの方々との協働作業を通じて理解が深まったこと, もっと勉強したいと思ったこと
  - 講演
  - SGD
  - 教育研究・教育心理について学ぶ clue がつかめた

- 班によってばらつきが大きい
- 市川先生の講演, その後の議論
- 教育心理学の話が参考になりました
- 市川先生の講演
- ワークショップは参加自体勉強になります. 不勉強なことがよく分かりました
- 実践研究の意義
- ワークショップと講演の両方があったこと
- 医学教育研究の必要性を再認識し, 本学部でも始める動機づけとなった, 市川先生の講演
- 教育心理学を学べた
- 他の機関, 分野の先生方と意見交換できたこと. 研究計画に必要なポイントがつかめたこと
- 研究全体の流れと難しさがわかった
- 方法論を考える良い機会となった
- 参加者の意識が高い
- 市川先生の講演を拝聴できたこと
- こういうWSが開かれるようになったという事実

6. 今回のワークショップで、芳しくなかった点を書いてください。

- もう少し十分なディスカッションをした方がいいのでは？(KJ法を使うなど)
- 個人的問題として統計学的素養がなかったので効果が低かったかも
- 会場が寒かった. 参加者の中に統計に関する知識に差がありすぎたこと(長所でもあるがロスも多い)
- 懇親会
- 統計解析手法の部分が不消化でもう少し時間をかけられるとさらに良かったと思います
- PC, メモリ
- グループ討論の目標の説明が不十分か？
- 医学研究で特別, 特殊な点を強調したWSであって欲しかった. 例えば, 対象が少人数であること, 交絡因子が多いこと, これらをどう克服するかが大切で, そのような点の配慮は少なかった. このWSで語られたものは臨床研究での研究デザインとほとんど変わらなかった. この点が期待はずれであった. ただし, 参加者の興味やレベルが色々なので, しょうがない面もある. 第1回なので仕方ないかとも思う
- パソコンの不備があった
- 配付資料が不足かな
- ディスカッション時間が短い
- 生物学的研究との違い
- ワークショップ発表に対する講師の先生方のコメントがもっと欲しかった. でなければやりっ放しになる.
- PCトラブルが多かった
- コンピューターの不備, 不都合でグループワークがあまり機能しなかった. 移動時間もロス
- 状況設定が少し難しかった
- グループの人数が多い
- コンピューター設置前に動作状況を確認しておいて欲しかった
- 仮想データは少し背景を説明してからの方がよかったかもしれない
- 1日で終わった方が好都合
- 統計実習は事前講義もなく, チューターもおらずかなりとまどった

7. その他気づいた点を書いてください。

- 継続して下さい
- 今回は総論として, 今後各論的に継続して下さい
- SGDにチューターが必要?
- 医学教育研究技法について, 資料, 情報を得られ有意義でした, 更なる発展を望みます
- PC持参でもよいか?
- 初日夕方寒かった
- 名札があると話しやすい
- デザイン, 統計に特化した方がよいかも(本来分けられないが)

- オーガナイズしてくださった先生方，スタッフに深謝申し上げます
- 続編を期待します。市川先生のレクチャーで，その実践研究（報告）レベルからでも書くことか  
らしか始まらないと感じました
- 懇親会の説明は参加者だけにした方がよかった
- PC等不具合が多かった

## ポストアンケートのまとめ

一定の評価は得られたものの、「まあまあよかった」という評価が全体的に目立つ印象である。時間配分を考えた際に，もう少し講演を短くし，グループディスカッションで十分時間がとれる形が望ましかったのであろう。今回のワークショップの続きがあれば是非参加したいという意見は多く，より上級な内容を期待している参加者が多いのだと思われる。

市川伸一先生の特別講演は自由記載コメントからも非常に評判が良かった。やはり，教育研究を実際に経験してきている教育学の専門家の協力は重要な課題である。また，よい講師を呼んでこのような講演を実施することで，参加費を取ってこういうワークショップをすることも受け入れられるのだと思われた。

今回のような意識の高い参加者を招いてワークショップを開くのであれば，より時間をたっぷりとって，例えば土曜日午前中から日曜日午後までといった形で，ディスカッションにもっと十分な時間がとれるようにすべきだろう。

## 医学教育研究に関連した参考図書

### 心理測定法

- Mehrens WA, Lehmann IJ. Measurement and evaluation in education and psychology (4th Ed.). Wadsworth. 1991 ——心理学, 教育学に関連した測定や評価に関して, 網羅的に記載しているが, 個々の分野での内容は適度な量にコントロールされている。
- Thorndike RM. Measurement and evaluation in psychology and education (6th Ed.). Prentice-Hall. 1997 ——古典的な心理測定学のテキストの1つ。内容は基本的なものが中心だが網羅的。
- Linn RL 編. 池田央ら編訳. 教育測定学原著第3版. C.S.L.学習評価研究所. 1992 ——教育測定学の全分野を網羅したテキスト。妥当性に関してMessickの定義を掲載している。
- 市川伸一編. 心理測定法への招待: 測定からみた心理学入門. サイエンス社. 1991 ——様々な領域における心理測定法について実例から初歩的な理論まで記載している。

### 研究の方法論やデザイン

- Fraenkel JR, Wallen NE. How to design & evaluate research in education (4th Ed.). McGrawhill. 2000 ——教育学的研究の基礎から応用までを実例やイラストを交えて分かりやすく解説している。
- Hulley SB, Cummings SR. 木原正博監訳. 医学的研究のデザイン: 研究の質を高める疫学的アプローチ. 医学書院 MYW. 1997 ——疫学的研究を中心に一般的な考え方を述べている。
- Pope C, Mays N. 大滝純司監訳. 質的研究実践ガイド: 保健・医療サービス向上のために. 医学書院. 2001 ——医療領域の質的研究に関して概説している。
- 舟島なをみ. 看護教育学研究: 発見・創造・証明の過程. 医学書院. 2002 ——特に研究に関連した主たるアイデアを質的にどのようにまとめていくべきかを詳しく述べている。
- 秋田喜代美, 恒吉僚子, 佐藤学編. 教育研究のメソドロジー: 学校参加型マインドへのいざない. 東京大学出版会. 2005 ——教育の場における実践型研究にこだわって教育研究を語っている。
- 南風原朝和, 市川伸一, 下山晴彦編. 心理学研究法入門: 調査・実験から実践まで. 東京大学出版会. 2001 ——心理学, 教育学領域で研究する際の標準的テキストと言える本。

### 評価(特にプログラム全体)

- Kern DEら著. 小泉俊三監訳. 医学教育プログラム開発: 6段階アプローチによる学習と評価の一体化. 篠原出版新社. 2003 ——医学教育領域のカリキュラム開発に関して実践的に記載している。特に評価については系統立っている。
- Worthen BR, Sanders JR, Fitzpatrick JL. Program evaluation: alternative approaches

and practical guidelines (2nd Ed). Longman. 1997 ——プログラム評価に関して考え方も含めて詳細に記載している。

#### 質問紙法

- 鎌原雅彦ら. 心理学マニュアル:質問紙法. 北大路書房. 1998 ——心理学, 教育学の研究に利用する質問紙の開発について基本的な内容を系統立てて記載している。
- 山本真理子編. 心理測定尺度集<1>:人間の内面を探る“自己・個人内過程”. サイエンス社. 2001
- 吉田富二雄編. 心理測定尺度集<2>:人間と社会のつながりをとらえる“対人関係・価値観”. サイエンス社. 2001
- 松井豊編. 心理測定尺度集<3>:心の健康をはかる“適応・臨床”. サイエンス社. 2001——いずれも妥当性検証が済んだ各種質問紙を厳選して掲載している。

#### 研究倫理

- Steneck NH.. 山崎茂明訳. ORI 研究倫理入門:責任ある研究者になるために. 丸善. 2005 ——研究者として果たすべき倫理的問題について概説している。

---

## 第1回医学教育研究技法ワークショップ報告書

---

2006年1月発行

編集及び発行 日本医学教育学会医学教育研究開発委員会  
委員長 東京大学医学教育国際協力研究センター 大西弘高  
〒113-0033 東京都文京区本郷 7-3-1 医学部総合中央館 212  
TEL: 03-5841-3534 FAX: 03-5802-1845  
E-mail: onishi-hirotaka@umin.ac.jp

印刷所 三鈴印刷株式会社  
〒101-0051 東京都千代田区神田神保町 2-32-1  
TEL: 03-5276-0811

---