

ラーニング・アウトカムのデジタル評価法

Digital Evaluation Method of Learning Outcome

胡 明（京都情報大学院大学）

Ming Hu (The Kyoto College of Graduate Studies for Informatics)

Abstract

In recent years, the attention to learning outcomes is increasing. The Entry/Exit test results of students from KCGI and the outside were investigated by using tools of Excel and R programming language. We analyzed learning outcomes of KCGI by comparing knowledge level discrepancy of before/after learning based on data from KCGI students only and the discrepancy between KCGI students and the outside students. Finally, we confirmed learning outcomes of KCGI by objective numerical visualization measurement method.

1. はじめに

近年、アメリカ、欧州、イギリスなど先進諸国はラーニング・アウトカム（学習成果）に対する関心を高めている。常に各国が目指すべき模範とされているアメリカの高等教育はラーニング・アウトカムを重視してきた。連邦教育省は各種のアクレディテーション機関すべてに各大学にラーニング・アウトカムを重視し、1989年に、明確に大学評価基準として、学生のラーニング・アウトカムの検証が設定されていることを条件とすることを決めた[1]。欧州では、1999年に始まり、2010年にヨーロッパとして高等教育圏を形成するため、「ボローニャ宣言」を声明した。そのため、行程化された「ボローニャ・プロセス」により、現在46カ国が参加し、アメリカの高等教育機関に匹敵できるように、ラーニング・アウトカムは欧州と各国の高等教育改革において、重要な役割を果たしている[1,2]。

近年、日本の少子化のため19歳～22歳の大学入学適齢の若者が年々減少している一方、大学進学率が高まり、「大学のユニバーサル」という望めば誰もが必ずどこかに進学できてしまう「大学全入時代」に入る時期を迎える。しかし、一方で企業や社会側は学生に求めるニーズ、学生への教育の質を高めている[3]。

その日本の現状と国際的な動向を背景に、ラーニ

ング・アウトカム（学習成果）の重要性が日本の高等教育界で取り上げられてきている。2008年、文部科学省が中央教育審議会で答申した「学士課程教育の構築に向けて」で、学生のラーニング・アウトカムの達成に向けた教育内容・方法を取り込む重要性和火急の対応すべき問題があると述べた[4]。なお、2012年8月に中央教育審議会から発表された「新たな未来を築くための大学教育の質的転換に向けて～生涯学び続け、主体的に考える力を育成する大学～」(答申)により、ラーニング・アウトカムの評価に当たっては、「学修時間の把握といった学修行動調査やアセスメント・テスト（学修到達度調査）、ルーブリック、学修ポートフォリオ等、どのような具体的な測定手法を用いたかを併せて明確にする」として示された[5]。

なお、ラーニング・アウトカムに対する重視度に比べて、実際にその状況と結果を把握するための評価指標に関する設定と分析も高めて行われないといけない。世界中、特に先進諸国の大学が様々な方法でラーニング・アウトカム評価を実施している。アメリカの南部のルイジアナ州に位置する私立大学であるDillard大学は大学の質を保証するため、「The Quality Enhancement Plan (QEP)」と呼ばれるプログラムを構築し、定期的にthe Collegiate Learning Assessment (CLA) など測定解析できるプログラム・アセスメントを実施し、5年間ですべ

での測定値を収集分析している [6]。

しかし、ほかの先進諸国に比べて、日本では、ラーニング・アウトカムの状況を把握するため、ラーニング・アウトカムの指標の設定とそれに関するアセスメントの実施とその結果の分析は十分に行われていたとは言えない。

葛城浩一 [7] は「在学中にどの程度能力が伸びたか」と、「専門分野の能力がどの水準にあるか」の2種類の主観的な判断に依拠するラーニング・アウトカム指標を設定し、日本全国の4年制大学18校56学部の1年生及び4年生（医学部・歯学部については6年生）を対象にラーニング・アウトカムを実施した。それに対して、客観的に測定可能なものを指標として、ラーニング・アウトカムを測定する場合はアメリカで多くの客観テストがあるのに対し、日本では一部の専門分野の範囲内でしか行われていないという現状がある。

2004年4月に開学した日本で最初のIT系の専門職大学院である京都情報大学院大学はこうした国際的な動向を背景に、アメリカのインタラクティブ社のサポートから最初の2004年度と2005年度に入学の学生に対して入学時と初年度終了時に下記科目のテストを行った。また、対象グループとして、KCGI入学生でない同レベルの学生グループ（Control）に対しても同じテストを実施した。

- ・一般ビジネス関連基礎知識
- ・IT 関連基礎知識
- ・インターネットビジネス関連基礎知識
- ・経済学関連基礎知識

その結果、カリキュラムに従って教育されたKCGI学生の1年間の成長ぶりは、上記のKCGIの4コアカリキュラム領域すべてにおいて、対象としたKCGI以外の学生に統計的に有意に勝っているという結果が得られた [8]。

その後、KCGIの作花一志とマリア IKEDA [9] はt検定と分散分析の解析方法を試みて、4個のExcel関数を駆使することにより、同じ結果を得ることができた。そして、長谷川明と作花一志 [10] はこのKCGIにおけるラーニング・アウトカムの測定について、2012年8月、近畿大学理工学部情報学科の第1回外部評価委員会において説明・報告を行った。

その解決方法の上、多重比較検定の解析方法も加えて、2010年10月から2016年10月までの入学学生及び2011年3月から2017年3月卒業学生の

テスト成績を分析した。まず、同じグループのKCGI学生が学習前と学習後にt検定と分散分析で各専門知識の差異があるかどうか、及びその差異の効果を効果量で検証した。さらに、多重比較法で各学期のKCGI学生のそれぞれの入口テストと出口テストで任意の2つのグループのデータの間に差異があるかどうかを検証し、箱ひげ図と平均推移図でそのデータの傾向を分析した。最後にt検定と効果量でKCGIの学生と他の学生を対象として、両グループ間に学習前と学習後の専門知識の差異があるかどうか、及びその差異の効果量を検証した。

2. データの収集とテスト内容

2017年2月上旬から3月末までに実施した以下のテスト結果データの集計を行った。3月末から4月中旬までにそのデータを分析した。

- ・KCGIの学生の2010年10月～2017年3月のデータ
- ・他の学生の2010年10月と2011年4月の入口テストデータと2011年3月と2013年3月の出口テストデータ

なお、2010年10月入学・2011年3月卒業～2013年10月入学・2015年9月卒業の時期にその4種類の基礎知識についてテストを行ったが、2014年4月入学・2016年3月卒業～2015年4月入学・2017年3月卒業の時期に下記の5種類のテストに変更した。

1. 一般ビジネス関連基礎知識
2. IT 関連基礎知識
3. インターネットビジネス関連基礎知識
4. 経営システム関連基礎知識
5. ネットワーク関連基礎知識

3. 分析方法

いくつかのグループの間に、統計的に平均値の違いがあるかどうかを判定する際に、グループの数により、よく使われる分析方法はt検定、z検定や分散分析（ANOVA）である。その中の2グループの間に平均値の違いがあるかどうかの検定は2群の平均値の差の検定という。すなわち、実験結果を使用前・使用後などの2グループ（群）に分け、それらの平均値の差が母集団においてもあるといってもよいかどうかを統計学的に検討する手法である

[11]。よく使われる検定手法はt検定とz検定である。

なお、t検定の3群以上への拡張とはよく言われる分散分析である。分散分析(Analysis of Variance, ANOVA)は3群以上の平均値の差についてF検定を実施することで、ある要因が実験結果に影響を与えたと言ってもよいかどうかを統計学的に検討する手法である[11]。

3-1 t検定

ある標本の平均値 $\bar{x}$ と母集団の平均値 μ が特定の値である μ_0 と差異があるかどうかを検定するには「 $\bar{x}=\mu_0$ 」という仮説(帰無仮説という)を設けて、統計量 $t=(\bar{x}-\mu_0)/\sqrt{s^2/n}$ に対する値が廃却域に入るかどうかで判定する。ただし、 n は標本数或は標本サイズであり、 s^2 は標本の不偏分散である。EXCELではvar関数から s^2 が求められる。

2群の平均値の差の検定について、標本サイズにより、最もよく使われる検定手法はt検定とz検定である。標本サイズ n が少数の場合はt分布(t検定)が用いられる。有意水準としては通常5%が採用される。すなわち、 $\bar{x}=\mu_0$ という事象が起こる確率 p (p 値)を計算し、それが0.05より小さければこの仮説を棄却し、両者に差異はないと判定するわけである。

ここではあるサンプルで有意水準5%のt検定を行う方法を述べる。分析した結果は図1のようなウィンドウが現れる。

	変数 1	変数 2
平均	32.9125	40.4625
分散	74.08085443	211.669462
観測数	80	80
ピアソン相関	-0.029897434	
仮説平均との差異	0	
自由度	79	
t	-3.943499948	
P(T<=t) 片側	8.63236E-05	
t 境界値 片側	1.664371409	
P(T<=t) 両側	0.000172647	
t 境界値 両側	1.99045021	

図1 t検定の分析結果

2つの方法によってこれら2群の平均値の差異がないという仮説の起こる確率は9.90066E-05と求まる。このような場合が起こる確率は1%未満であり、有意水準5%を下回り、当初の仮説は棄却され

る。すなわち、めったに起こらないことであり、言い換えれば2群の間には平均値の差異が認められるのだが「2群の平均値に差異が認められないという仮説は棄却される」という消極的な言い方をする。

この方法で標本サイズが少ない場合には、KCGIグループの入口テストと出口テストの平均値に差異が認められるかどうかが判定される。

3-2 分散分析

分散分析はその名の通り、実験や調査から得られた群間のデータの平均値の差があるかどうかをそのデータの分散を分析する統計的な手法である。要因の数によって、一元配置分散分析(1要因の場合)、二元配置分散分析(2要因の場合)、多元配置分散分析(3要因以上の場合)と区別して呼ばれる。一元配置分散分析に用いる検定統計量であるF値は次の式のように2つの分散の比で定義される[11]。

$$F = \frac{\text{目的となる要因効果(群間変動)の分散}}{\text{誤差効果(群内変動)の分散}}$$

実際に分析するとき(2群の場合)、EXCELの分析ツールから行う。分析した結果は図2のようなウィンドウが現れる。

分散分析: 一元配置						
概要						
グループ	標本数	合計	平均	分散		
入口	80	2633	32.9125	74.080854		
出口	80	3237	40.4625	211.66946		
分散分析表						
変動要因	変動	自由度	分散	帰された分散	P-値	F 境界値
グループ間	2280.1	1	2280.1	15.958687	9.90066E-05	3.90088696
グループ内	22574.275	158	142.8751582			
合計	24854.375	159				

図2 分散分析の分析結果

これによってこれら2群の平均値の差異がないという仮説の起こる確率は9.90066E-05と求まる。このような場合が起こる確率は1%未満であり、有意水準5%を下回り、当初の仮説は棄却される。

この方法で標本サイズが大きい場合には、KCGIグループの入口テストと出口テストの平均値に差異が認められるかどうか判定される。そして、入口テスト或は出口テスト内に各専門知識の間にテスト結果平均値の差があるかどうか判定できる。

3-3 多重比較法

分散分析で複数の群の間にデータの平均値の差があるかどうか判定できたが、どの群とどの群との間に差があるかがまだ分からない。そのため、2群

に対し、繰り返してt検定を行うと、多重性という大きな問題が発生してしまうのである。もう一方、群の数を増やすと、帰無仮説が簡単に棄却域に入りやすくなってしまう問題もある。それらの問題を避けるために、多重比較法（或は多重比較検定）という統計検定の手法がある。多重比較法は3つ以上の群において、どの群とどの群との間に差があるのか測定する手法の総称である [11]。

いろいろな多重比較法において、よく使われる検定方法はTukey-Kramer（各群の分散が等しい場合）と Games-Howell（各群の分散が等しくない場合）という手法である。Tukey-Kramer 法の統計量である q 値は次の式のように定義される。

$$q_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{V_w \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}}$$

ただし、 $\bar{x}_i$ と $\bar{x}_j$ は第 i 群と第 j 群の平均値であり、 n_i と n_j は第 i 群と第 j 群の標本サイズであり、 V_w は不偏誤差分散である。Games-Howell の場合は次の式のように定義される。

$$q_{ij} = \frac{|\bar{x}_i - \bar{x}_j|}{\sqrt{\frac{U_i}{n_i} + \frac{U_j}{n_j}}}$$

ただし、 U_i と U_j は第 i 群と第 j 群の標本標準偏差である。比較する 2 群の間に平均値の差があるかどうか検定するとき、その q 値とスチューデント化された範囲の q 分布の値の大小関係や p 値は有意水準 5% 以下になるかどうかで判断する。

Excel で単一の関数や分析ツールで多重比較法の分析は無理だが、R というフリーソフトで分析できる。R は統計解析用に開発された有名なフリーソフトで、統計学を基本から学び多変量解析までを修得するのに役立ち、ビジネス界で実用されている。またグラフィックス機能が充実していて 3D 描画も容易にできるので、シミュレーションに適している。そのため近年データサイエンスのツールとして注目されている。

R は非常にたくさんの関数、ライブラリを持ち、しかも毎月のように増補され、適用範囲は多方面に広がっている [12]。

実際に分析するとき、R に用意されている TukeyHSD() 関数を使うのか、或は次の図 3 に表示された青木 繁伸から提供するコードパッケージで簡単に求める [13, 14]。

```

R E:\00 研究資料\00 教育効果テスト\tukey_test.R - RIDE
y <- read.table("tukey2.txt",header=TRUE) #データを読み込む。
str(y) #Rの内部のデータフレームyの2つ変数の内容を表示確認する。
by(y$成績, y$学期別, summary) #3つの群の基本統計量を計算する。
boxplot(y$成績~y$学期別) #3つの群の箱ひげ図を描く。
#library(Rcmdr)
plotMeans(y$成績, y$学期別, error.bars="se") #平均値の推移図を描く。
source("http://aoki2.si.gunma-u.ac.jp/R/src/tukey.R", encoding="euc-jp")
#青木先生のコードパッケージを呼び付ける。
#tukey(y$成績, y$学期別) #Tukey-Kramer法の多重比較法
tukey(y$成績, y$学期別, method="G") #Games-Howell法の多重比較法

```

図 3 多重比較法の R コード

分析した結果は図 4 - 図 7 のようなウィンドウが現れる。

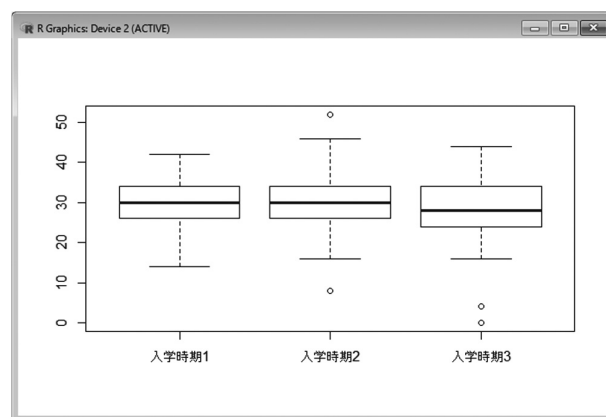


図 4 3つの群の箱ひげ図

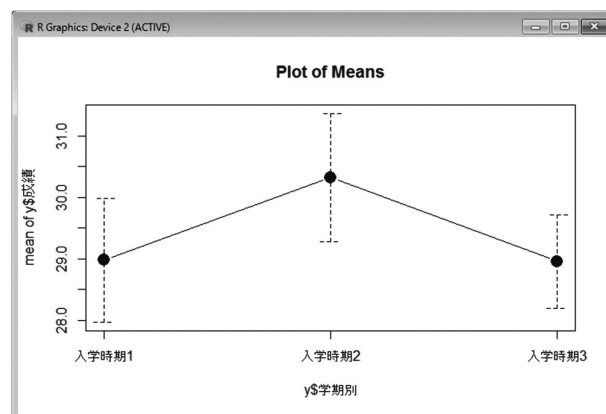


図 5 3つの群の平均値の推移図

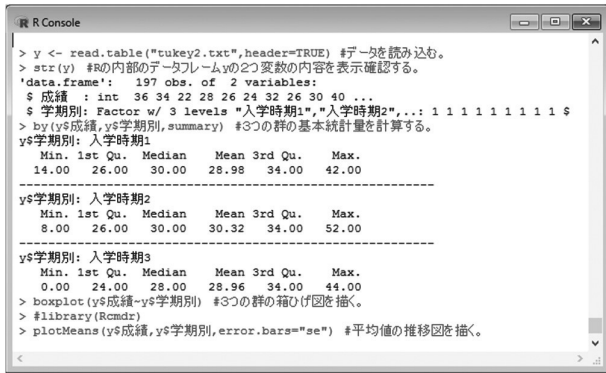


図6 3つの群の基本統計量



図7 Games-Howell 法の多重比較法

図7の[`$Games-Howell`]の欄にp値を見ると、優位水準5%以下のp値がないので、どの水準間にも有意差のあるテスト成績はないということが分かった。

3-4 検定分析の効果量

ここまで、t検定や分散分析(ANOVA)などいろいろな手法で複数の群の間に平均値の差があるかどうかを統計的に検討した。そのため、p値と呼ばれているよく使われる確率が検定分析において、最も重要な指標である。しかし、[15]に示したように、「p値は標本数が多くなればなるほど、実質的な差がなかった場合でも、p値は小さくなり、統計的に優位であるという結果が得られやすくなるという大きな問題を持っている。」

効果量とは、「効果の大きさ」を表す標準化された指標である。それを表す指標の数が多いが、t検定などで2つの群の間に平均値の差についての検定は、次の式のように定義された代表的な指標であるCohen's dである[16]。

$$d = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{U_i^2 + U_j^2}{2}}}, \quad (\text{各群の標本数が同じ場合})$$

$$d = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\frac{(n_i - 1)U_i^2 + (n_j - 1)U_j^2}{n_i + n_j}}}$$

ただし、 $\bar{x}_i$ と $\bar{x}_j$ は第i群と第j群の平均値であり、 n_i と n_j は第i群と第j群の標本サイズであり、 U_i と U_j は第i群と第j群の標本標準偏差である。もう一方、分散分析のとき、次の式のように定義される相関比 η の2乗 η^2 値(eta squared, イータ2乗値)がよく使われる[16]。

$$\eta^2 = \frac{\text{ある要因の平方和 (SS}_{effect})}{\text{全体の平方和 (SS}_{total})}$$

d値は1を超えるのが可能だが、 η^2 値は0から1までの値を取る。どちらの指標でも、絶対値が大きさと効果の大きさが一致している。

η^2 値の算出について、具体的に次のような手順で行う[2]。

$$SS_{effect} = \sum (\text{各群の平均値} - \text{全体の平均値})^2 \times \text{各群の標本数}$$

$$SS_{total} = \sum (\text{各データ} - \text{全体の平均値})^2 = \text{全体の平均に対する偏差の平方和}$$

そして、偏差平方和はExcelの関数DEVSQより求めることができる。

効果量によって、効果の大きさの判断は以下の表1で求めたものである[15,16]。

表1 検定(分析)の種類ごとに見る代表的な効果量の指標(一部)と大きさの目安

使用される検定(分析)	効果量の指標	効果量の目安		
		小	中	大
t検定	d値	0.20	0.50	0.80
一元配置分散分析	η^2 値	0.01	0.06	0.14

4. 結果と考察

4-1 KCGI 学生の入口・出口テストデータの比較分析結果

まず、前半に紹介した t 検定の p 値と一元配置分散分析の p 値及び効果量の指標 d 値と η^2 値を用いて、各学期の KCGI 学生の入口・出口テストデータの差があるかどうか、およびその差の大きさを KCGI の教育効果について分析する。ここで、4 種類の 2010 年 10 月入学・2012 年 9 月卒業のデータと 5 種類の 2014 年 10 月入学・2016 年 9 月卒業のデータを例として利用する。

表 2 2010 年 10 月入学・2012 年 9 月卒業 (4 種類)

テストの問題	テストの群	平均値	t 検定		分散分析	
			p 値	d 値	p 値	η^2 値
一般ビジネス関連	入口	28.98	0.02	0.49	0.03	0.06
	出口	43.50				
IT 関連	入口	40.74	0.00	1.03	0.00	0.22
	出口	78.69				
インターネットビジネス関連	入口	38.16	0.00	0.94	0.00	0.19
	出口	48.33				
経済・経営学関連	入口	24.39	0.00	0.63	0.01	0.09
	出口	30.90				

2010 年 10 月入学・2012 年 9 月卒業の比較結果を見ると、4 種類のテストについて、すべての出口テスト結果が優れており、この違いは統計的に有意差があること、そして、「一般ビジネス関連」を除く、ほかの 3 種類のテスト結果は大きい効果が出ることが分かった。

表 3 2014 年 10 月入学・2016 年 9 月卒業 (5 種類)

テストの問題	テストの群	平均値	t 検定		分散分析	
			p 値	d 値	p 値	η^2 値
一般ビジネス関連	入口	32.91	0.00	0.63	0.00	0.09
	出口	40.46				
インターネットビジネス関連	入口	44.73	0.00	0.89	0.00	0.17
	出口	56.48				
経営システム関連	入口	26.46	0.00	0.61	0.00	0.09
	出口	53.98				
IT 関連	入口	42.93	0.00	1.23	0.00	0.28
	出口	59.95				
ネットワーク関連	入口	24.24	0.00	0.95	0.00	0.187
	出口	43.80				

2014 年 10 月入学・2016 年 9 月卒業の比較結果を見ると、5 種類のテストについて、すべての出口テスト結果が優れており、この違いは統計的に有意差があること、そして、「一般ビジネス関連」と「経営システム関連」を除く、ほかの 3 種類のテスト結果は大きい効果が出ることが分かった。

4-2 KCGI 入口テストの分析結果

ここで分野別で箱ひげ図によって、「一般ビジネス関連基礎知識」を例として、各入学時期の入口テスト結果のデータ分布を示し、その平均値の推移を平均推移図で表す。最後、前半に説明した多重比較法によって、各入学時期の入口テストの比較結果(一部)を判定する。(有意差があった場合はグレイ塗り)

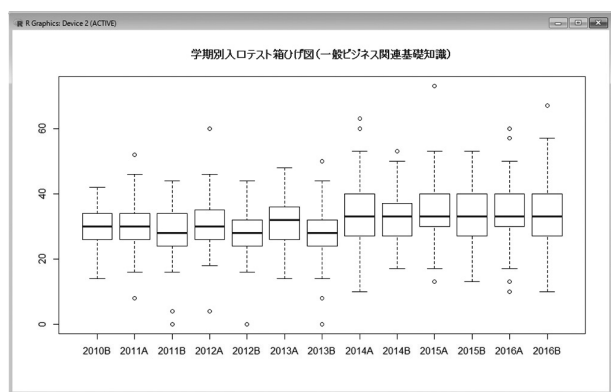


図8 入口テスト箱ひげ図

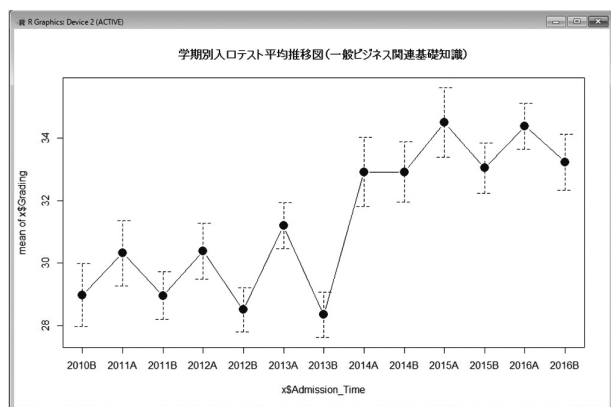


図9 入口テスト平均推移図

「一般ビジネス関連基礎知識」各学期の入口テスト結果の箱ひげ図（図8）を見ると、2010年10月（2010 B）から2013年10月（2013 B）までの入学生の一一般ビジネス関連基礎のバラツキと比べ、2014年4月（2014 B）から2016年10月（2016 B）までの入学生の「一般ビジネス関連基礎知識」のバラツキのほう広がっているのは分かった。なお、2010年10月（2010 B）から2013年10月（2013 B）までの入学生の中に「一般ビジネス関連基礎知識」は非常に足りない場合は多いものの、2014年4月（2014 B）から2016年10月（2016 B）までの入学生の中に、非常に豊かな「一般ビジネス関連基礎知識」を持っている学生の数が多いことが分かった。

なお、「一般ビジネス関連基礎知識」各学期の入口テスト結果の平均推移図（図9）を見ると、2014年4月までの入学生より、2014年4月以降の入学生は豊かな「一般ビジネス関連基礎知識」を持っていることが分かった。そして、全体に見ると、10月の入学生の成績より、4月の入学生の成績が良かったことも明らかに示された。これは多分10月

の入学生はほとんど留学生と違い、4月の入学生に京都コンピュータ学院（KCG）など日本国内から学生もいるのは一つの原因だと想定できる。

群間 ^a	p 値 ^a	群間 ^a	p 値 ^a	群間 ^a	p 値 ^a
2010B-2011A ^a	1.000 ^a	2011B-2013B ^a	1.000 ^a	2013A-2014B ^a	0.036 ^a
2010B-2011B ^a	0.999 ^a	2011B-2014A ^a	0.832 ^a	2013A-2015A ^a	0.001 ^a
2010B-2012A ^a	1.000 ^a	2011B-2014B ^a	0.887 ^a	2013A-2015B ^a	0.001 ^a
2010B-2012B ^a	0.998 ^a	2011B-2015A ^a	0.685 ^a	2013A-2016A ^a	0.002 ^a
2010B-2013A ^a	1.000 ^a	2011B-2015B ^a	0.247 ^a	2013A-2016B ^a	0.000 ^a
2010B-2013B ^a	0.850 ^a	2011B-2016A ^a	0.647 ^a	2013B-2014A ^a	0.970 ^a
2010B-2014A ^a	0.206 ^a	2011B-2016B ^a	0.085 ^a	2013B-2014B ^a	0.965 ^a
2010B-2014B ^a	0.299 ^a	2012A-2012B ^a	0.916 ^a	2013B-2015A ^a	0.893 ^a
2010B-2015A ^a	0.095 ^a	2012A-2013A ^a	1.000 ^a	2013B-2015B ^a	0.401 ^a
2010B-2015B ^a	0.021 ^a	2012A-2013B ^a	0.292 ^a	2013B-2016A ^a	0.864 ^a
2010B-2016A ^a	0.091 ^a	2012A-2014A ^a	0.019 ^a	2013B-2016B ^a	0.106 ^a
2010B-2016B ^a	0.002 ^a	2012A-2014B ^a	0.050 ^a	2014A-2014B ^a	1.000 ^a
2011A-2011B ^a	0.998 ^a	2012A-2015A ^a	0.002 ^a	2014A-2015A ^a	1.000 ^a
2011A-2012A ^a	1.000 ^a	2012A-2015B ^a	0.001 ^a	2014A-2015B ^a	0.997 ^a
2011A-2012B ^a	0.992 ^a	2012A-2016A ^a	0.003 ^a	2014A-2016A ^a	1.000 ^a
2011A-2013A ^a	1.000 ^a	2012A-2016B ^a	0.000 ^a	2014A-2016B ^a	0.992 ^a
2011A-2013B ^a	0.646 ^a	2012B-2013A ^a	0.862 ^a	2014B-2015A ^a	1.000 ^a

図10 入口テスト学期別有意差の検定

ここで、多重比較法によって、各入学時期の入口テストの比較結果（一部）を見ると（図10）、「一般ビジネス関連基礎知識」について、2011年4月、2012年4月（及び2013年4月）のデータとほかのデータと比較し、有意差が出る場合は多いが、全体に見ると、多数の入学学期別の入口テスト結果の間に有意差がなかったことが分かった。

4-3 KCGI 出口テストの分析結果

ここで分野別で箱ひげ図によって、「一般ビジネス関連基礎知識」を例として、各入学時期の出口テスト結果のデータ分布を示し、その平均値の推移は平均推移図で表す。最後、前半に説明した多重比較法によって、各入学時期の出口テストの比較結果（一部）を判定する。（有意差があった場合はグレイ塗り）

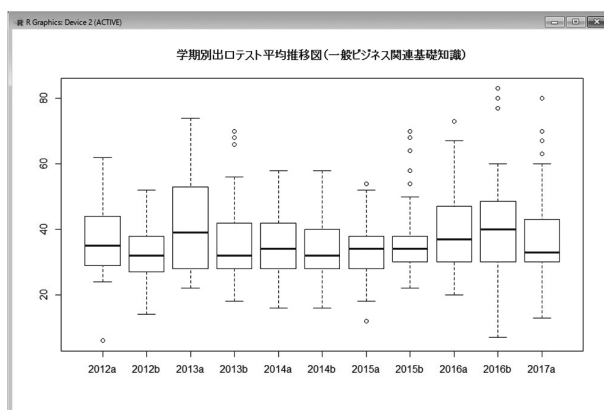


図11 出口テスト箱ひげ図

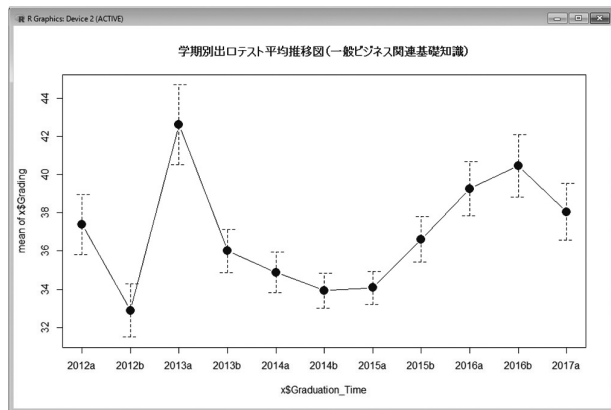


図 12 出口テスト平均推移図

「一般ビジネス関連基礎知識」各学期の出口テスト結果の箱ひげ図（図 11）を見ると、まず入口テスト結果のように、非常に低い成績の場合は少なくなった。逆に、知識が豊かな学生の数、特に近年、増加している傾向が見える。

なお、「一般ビジネス関連基礎知識」の入口テスト結果と違う、「一般ビジネス関連基礎知識」各学期の出口テスト結果の平均推移図（図 12）を見ると、10 月入学生の成績と 4 月の入学生の成績と比べて、明確な差がなくなる、そして、入学時期にかかわらず、学期別成績が増加している場合もあった（例えば、2014 年 10 月から 2016 年 10 月まで）。

群間 ^a	p 値 ^a	群間 ^a	p 値 ^a	群間 ^a	p 値 ^a
2012a2012b ^a	0.550 ^a	2013a2013b ^a	0.180 ^a	2014a2016a ^a	0.327 ^a
2012a2013a ^a	0.651 ^a	2013a2014a ^a	0.052 ^a	2014a2016b ^a	0.145 ^a
2012a2013b ^a	1.000 ^a	2013a2014b ^a	0.012 ^a	2014a2017a ^a	0.814 ^a
2012a2014a ^a	0.963 ^a	2013a2015a ^a	0.014 ^a	2014b2015a ^a	1.000 ^a
2012a2014b ^a	0.717 ^a	2013a2015b ^a	0.323 ^a	2014b2015b ^a	0.786 ^a
2012a2015a ^a	0.751 ^a	2013a2016a ^a	0.962 ^a	2014b2016a ^a	0.067 ^a
2012a2015b ^a	1.000 ^a	2013a2016b ^a	0.999 ^a	2014b2016b ^a	0.026 ^a
2012a2016a ^a	0.998 ^a	2013a2017a ^a	0.789 ^a	2014b2017a ^a	0.397 ^a
2012a2016b ^a	0.955 ^a	2013b2014a ^a	1.000 ^a	2015a2015b ^a	0.821 ^a
2012a2017a ^a	1.000 ^a	2013b2014b ^a	0.939 ^a	2015a2016a ^a	0.074 ^a
2012b2013a ^a	0.009 ^a	2013b2015a ^a	0.957 ^a	2015a2016b ^a	0.029 ^a
2012b2013b ^a	0.808 ^a	2013b2015b ^a	1.000 ^a	2015a2017a ^a	0.430 ^a
2012b2014a ^a	0.987 ^a	2013b2016a ^a	0.771 ^a	2015b2016a ^a	0.938 ^a
2012b2014b ^a	1.000 ^a	2013b2016b ^a	0.466 ^a	2015b2016b ^a	0.710 ^a
2012b2015a ^a	1.000 ^a	2013b2017a ^a	0.999 ^a	2015b2017a ^a	1.000 ^a

図 13 出口テスト学期別有意差の検定

ここで、多重比較法によって、各卒業時期の出口テストの比較結果（一部）を見ると（図 13）、「一般ビジネス関連基礎知識」について、いくつか有意差が出る場合は多いが、全体に見ると、2013 年 9 月以降の卒業時期別の出口テスト結果の間に有意差がなかったことが分かった。

4-4 KCGI の学生と他の学生の入口・出口テストの分析結果

同じグループの学生の入口・出口テストは入口より出口の方の点数が上回っているのは普通だったが、ここで、2011 年 4 月入学の KCGI の学生（I）と他の学生（II）との 4 種類の入口・出口テストの結果を t 検定で分析し、ラーニング・アウトカムの結果を示す。

表 4 2011 年 4 月入学の KCGI の学生と他の学生の入口テスト

入口	一般ビジネス関連	IT 関連	インターネットビジネス関連	経済経営関連
I 平均	30.32	53.11	45.34	29.43
II 平均	41.14	72.86	56.86	50.43
p 値	0.03	0.00	0.00	0.00
d 値	-1.36	-1.36	-1.05	-1.60

表 5 2011 年 4 月入学の KCGI の学生と他の学生の出口テスト

入口	一般ビジネス関連	IT 関連	インターネットビジネス関連	経済経営関連
I 平均	42.61	65.43	53.00	38.77
II 平均	37.14	74.29	54.00	43.14
p 値	0.17	0.07	0.82	0.37
d 値	—	—	—	—

2011 年 4 月入学の KCGI の学生と他の学生の入口・出口テスト結果を見ると（表 4、表 5）、学習前に他の学生の成績は全体的に KCGI の学生より良かったし、p 値より顕著な差異があるのは示された。しかし、学習の後、両者の差異が認められなかった。なお、一般ビジネス関連知識と、平均的に、KCGI の学生は他の学生の成績を上回っていることも分かった。

5. おわりに

上記, Excel と R 言語の二つの統計解析ツールを用いて, t 検定, 分散分析, 多重比較法などの統計手法で 2010 年 10 月～2017 年 3 月の間の KCGI の学生及び 2011 年 4 月・2013 年 3 月の他の学生が行った入口・出口テストを分析した。同じグループの学生が学習前と学習後に知識レベルの差異があるかどうか, 及びその差異の効果を検証した。なお, KCGI の学生と他の学生を対象として, 両グループ間に学習前と比べて, 学習後知識レベルの差異があるかどうかを分析した。ここで, 客観的に測定可能なものを指標として, 数値的にラーニング・アウトカムとその分析を行った。さらに, 統計ツールで分析結果を視覚表現化した。この結果で, 教育の質を高め, ラーニング・アウトカム計測が今後もっと重要な役立ちになれると思う。

謝辞

京都情報大学院大学の作花一志教授に様々なアドバイスをいただき, 筆者を導いていただいております。深く感謝いたします。

【参考文献】

- [1] 川嶋太津夫, “ラーニング・アウトカムズを重視した大学教育改革の国際的動向と我が国への示唆”『名古屋高度教育研究』, 第 8 号, 173-191
- [2] 館 昭, “ボローニャ・プロセスの意義に関する考察: ヨーロッパ高等教育圏形成プロセスの提起するもの”, 2010
- [3] 社団法人関西経済同友会, “社会が求める大学の人材輩出戦略 ～まずは学部教授会の改革から～”, 2009
- [4] 中央教育審議会, “学士課程教育の構築に向けて (答申)”, 2008
- [5] 中央教育審議会, “新たな未来を築くための大学教育の質的転換に向けて ～生涯学び続け, 主体的に考える力を育成する大学～ (答申)”, 2012
- [6] 御厨 まり子, “学生のラーニング・アウトカムの取組みについて ～米国 Dillard University の調査をもとに～”, 明星大学明星教育センター研究紀要, 5 (2015): 143-150
- [7] 葛城 浩一, “アウトカム指標のあり方を考える”, 大学論集, 41 (2010): 439-454
- [8] デール・マン, 2004 年度—2005 年度入学対象学生結果分析のレポート, Interactive 社専務理事, 2006。
- [9] マリア IKEDA, 作花 一志, 学習効果のデジタル評価法, 2008
- [10] 長谷川 明, 作花 一志, ラーニング・アウトカムの測定, NAIS Journal, 2013
- [11] 栗原 伸一, 入門統計学 ～検定から多変量解析・実験計画法まで～, オーム社, 2014
- [12] 作花 一志, 胡 明, R による数値計算と統計解析, NAIS Journal, 2014
- [13] 青木 繁伸, テューキーの方法による多重比較, <http://aoki2.si.gunma-u.ac.jp/R/tukey.html>
- [14] 梶山 喜一郎, コピペで学ぶ R でテクニカルデータプレゼンテーション, http://monge.tec.fukuoka.ac.jp/r_analysis/Or_analysis.html#anova
- [15] 水本 篤, 竹内 理, 効果量と検定力分析入門—統計的検定を正しくつかるために—, 2010 年度部会報告論集「より良い外国語教育のための方法」, pp.47-73, 2011 年
- [16] 水本 篤, 竹内 理, 研究論文における効果量の報告のために—基礎的概念と注意点—, 関西英語教育学会紀要英語教育報告, 31, pp.57-66, 2008

◆著者紹介

胡 明 Ming Hu

京都情報大学院大学講師。

京都大学大学院情報学研究科数理工学専攻博士課程修了。情報学博士。研究分野はナッシュ均衡, マルチリーダー・フォロワゲームと均衡制約付き均衡問題。元日本学術振興会特別研究員 (PD2)。元中国江蘇科技大学講師。