

第Ⅲ章 データに基づく総合的検証

本章では、3期にわたる評価の結果について、統計的な分析手法を用いて可視化することによって、今後の評価方法の改善に向けた方法論的整備の基礎的資料を提示する。第Ⅲ章第1節では、中期目標の達成状況評価における教育及び研究の大項目の判定結果に関して、第1期から第3期までの全体的状況の変化を分析し、その変化の要因について、評価方法の変更の影響を含めて検討した。第2節では、第1節と同様に法人の学部・研究科等を単位とする現況分析の結果について分析したが、現況分析が、学部・研究科等の現況分析単位（以下、必要に応じて「学部・研究科等」「学部等」及び「研究科等」という用語を用いる。）について学系ごとに教育及び研究の活動と成果の状況の評価する点を考慮して検討を加えた。第3節では、第3期の研究業績水準判定において複数の評価者が同一の業績を評価した結果のデータを分析し、評価結果の信頼性を検討した。この分析で使用したデータは公表されたものではないが、研究の成果に関する分析項目の現況分析において評価者が判断の根拠の一つとしている。第4節では、国立大学法人等の教育研究の評価において定量的な指標を活用する手法である「データ分析集」における指標と評価結果との関係について考察した。

本章における検討は、評価者や法人等からの意見ではなく、評価結果を統計的に分析することによって、評価方法の改善、改良を図るための基礎を提供するためのものである。特に第1期から第3期への変化を検討する場合には、評価基準、評価方法の変更を考慮する必要がある。この検討結果がただちに国立大学法人等における教育研究の状況とその変化を表わすものではないことに強く留意する必要がある。また、各期間の比較を行うために実際の評価結果を本分析のために設定して点数に置き換えるなどの処理を行っている点、また、それゆえに、個別の法人、学部・研究科等ではなく、評価結果の全体的状況に関する検討である点についても、本章の検討結果を理解する際に留意しなければならない。各節末の考察の内容は、以上の点に留意して行った検討の状況を示すものであり、最終的な検討結果ではない。

本章で利用した統計的分析手法については、巻末に付した「本検証における統計手法・用語について」を参照されたい。

1. 中期目標の達成状況評価に関する分析

1.1 教育・研究の大項目の判定結果に関する第3期までの比較

第1期から第3期までの法人の教育・研究の大項目の判定結果を対象とし、全体としてどのような変化が見られたかについて分析を行った。

1.1.1 方法

第1期は平成22年度に行われた評価、第2期は平成28年度に行われた評価、第3期は令和2年度に実施された4年目終了時評価のデータを用いた。分析対象とした法人の総数は、90法人であった¹。データは全て機構のウェブサイトにて公表されており²、それらを基に各法人の評価結果から、判定を示す記述を基に点数化を行った。比較分析をするに当たり、第1期、第2期及び第3期で判定基準が異なるため、評価点2を標準と定め、1～4の評価点として定めた(表3-1-1)。

表3-1-1 比較のために統一した評価点と判定基準

評価点	判定基準
4	標準より2段階上回る
3	標準より1段階上回る
2	標準
1	標準より1段階下回る

表3-1-2 各期の判定基準に基づいた評価点への点数化

第1期および第2期			第3期		
点数化	判定を示す記述	備考	点数化	判定を示す記述	備考
4	非常に優れている		4	特筆すべき進捗	
3	良好		3	計画以上の進捗	
2	おおむね良好(標準)		2	順調(標準)	
1	不十分		1	おおむね順調	
0	重大な改善事項	該当なし	0	遅れている	該当なし
			0	重大な改善事項	該当なし

これに基づき、標準点が2となるように各期の評価点を算出した。以上の統一した評価点と第1期から第3期の判定を示す記述との関係について、表3-1-2に示す。第1期及び第2期は「おおむね良好」を標準とした(表3-1-2左)。第3期は「順調」を標準として補正した(表3-1-2右)。なお、補正後0以下の法人は存在しなかったため(備考欄で「該当なし」)、分析には評価点1～4の4段階を用いた。

以上の点数化された評価結果(評価点)を用いて、教育・研究の大項目(2水準)及び第1期～第3期(3水準)の2つの要因の影響について、二元配置分散分析(対応あり)を適用することで検討を行った。

¹ 筑波大学の第3期のデータは、大項目の構成が異なっていたため、分析には用いなかった。

² https://www.niad.ac.jp/evaluation/research_evaluation/kokuritikyoudou/

1.1.2 結果と考察

教育・研究の大項目の評価点の平均値の変化を示す（図 3-1-1）。

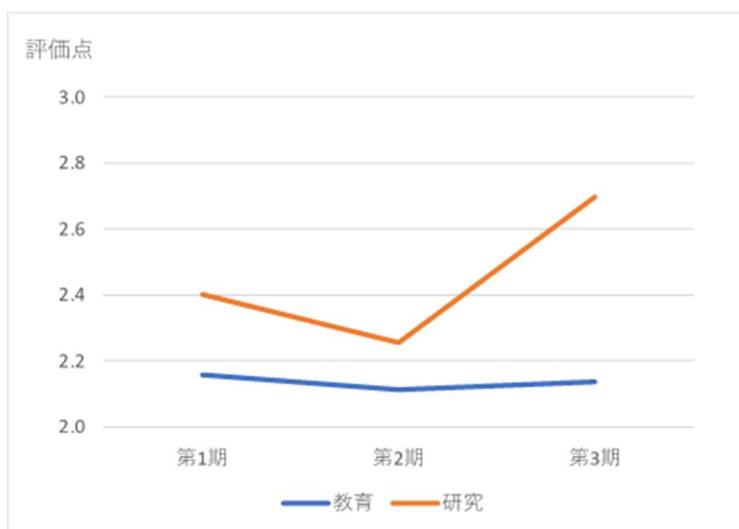


図 3-1-1 教育・研究の大項目の評価点の平均値の変化

分散分析の結果は

- ・ 教育・研究の違いによる主効果が有意 $F(1, 88) = 46.4, p < .01, \eta_p^2 = .345$
 - ・ 中期目標期間の違いによる主効果が有意 $F(2, 159) = 6.93, p < .01, \eta_p^2 = .073$
 - ・ 交互作用が有意 $F(2, 176) = 11.1, p < .01, \eta_p^2 = .112$
- 交互作用が有意であったので、単純主効果の検定を行ったところ、
- ・ 教育に関して、中期目標期間の違いによる主効果が有意差なし $F(2, 155) = .238, p = .76, \eta_p^2 = .003$
 - ・ 研究に関して、中期目標期間の違いによる主効果が有意 $F(2, 162) = 12.9, p < .01, \eta_p^2 = .128$ 、Bonferroni 法による多重比較³の結果、第 1 期と第 3 期、第 2 期と第 3 期の間で有意差が見られた。 $(p < .01)$
 - ・ 第 1 期 $F(1, 89) = 14.7, p < .01, \eta_p^2 = .142$ 、第 2 期 $F(1, 89) = 7.91, p < .01, \eta_p^2 = .082$ 、第 3 期 $F(1, 88) = 38.7, p < .01, \eta_p^2 = .305$ 、の全てにおいて、研究の方が教育よりも有意に高かった。

以上をまとめると、研究に関する評価点の方が、中期目標期間の違いに関わらず教育に関する評価点よりも高く、教育に関する評価点は中期目標期間の違いであまり変化していないことがわかり、研究に関する評価点に関しては特に第 3 期で高くなっていることが統計的に確認された。

³ 以降、多重比較には Bonferroni 法を用いた。

研究に関する評価点の方が総じて教育に関する評価点よりも高い傾向は、後述する現況分析においても同様に見られる。達成状況評価と現況分析は、別の概念に基づく評価であるため、教育と研究それぞれの切り口で比較することはあまり意味を持たないが、研究の方が評価点が高いということが一般的な傾向だとすると、大学評価においては総じて研究の方が高く評価されやすいことを示しているかもしれない。

また、第3期の研究の評価点が高くなった要因として、法人の研究の質が向上していること、評価者が高く評価するようになったこと、評価方法が変化したこと、など複数の可能性が考えられるが、検証に用いたデータのみからでは判断できない。

評価方法の変更による可能性としては、今回用いた第3期のデータが第1期及び第2期と違い4年目終了時の評価結果であることや、表3-1-2に示したように、判定基準が第3期では変更されていることなどが影響している可能性もある。次項では、これらの評価方法の変更による影響について分析する。

1.2 評価方法の変更による影響

第3期の評価では、中期目標の達成状況評価に関し、大別して2つの変更を行った⁴。

1つは、中期計画を3段階判定、中期目標の小項目を5段階判定とし、法人ごとに具体的な中期目標が設定されている小項目の評価を重視した仕組みとした。対して、第2期までは、中期計画・小項目ともに4段階判定とし、中期計画の段階判定結果を積み上げて、大項目の段階判定結果を算出した。このことが、第2期までの大項目判定結果が2判定「おおむね良好（標準）」に集中した一因として挙げられる。判定結果が2判定に集中したことは、大学間の達成状況の違いを判別できない評価であり、判別を意図的に回避しているのではないかとの憶測を招いたため、第3期では中期計画より一段階上の中期目標の小項目の段階判定結果を大項目に積み上げること、小項目の判定段階も1つ増やし5段階とし、達成状況の違いを明瞭に判別できるように評価方法を設計した。

もう1つの変更として、教育・研究の現況分析結果を教育・研究の大項目判定結果にそれぞれ加算・減算し、達成状況評価に反映させることとした。これは、現況分析結果をより有効に活用することや、1つ目の変更の理由と同様に、判定結果が2判定に集中することに対する社会からの批判に応える意味もあった。

ここでは、後者の変更について、現況分析結果の加算・減算前後の達成状況評価の結果を比較することにより、その影響がどの程度であったかを検証する。

1.2.1 方法

前項に示したものと同様のデータを用いて、加算・減算前後の評価点を比較した。前項の分析結果を踏まえ、第2期からの評価点が加算・減算前後でどう変化しているかにも着目し、グラフ表現により可視化することで比較を行った。

1.2.2 結果と考察

教育の達成状況評価結果への現況分析結果の加算・減算の影響について、第2期からの評価点の平均値の変化を図3-1-2に示す。これを見ると、第3期において加算・減算前は評価点が低下しているが、加算・減算後にはほぼ第2期と同じになっていることが読み取れる。

⁴ 詳細については、「参考資料2，文部科学省国立大学法人評価委員会総会（第67回）R3.6.30」を参照のこと。 https://www.mext.go.jp/kaigisiryoy/content/20210624-mxt_hojinka-000016269_27.pdf

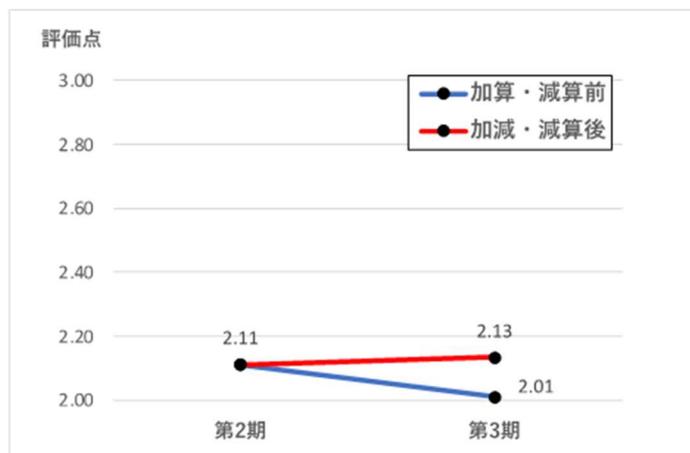


図 3-1-2 教育の達成状況評価の結果への現況分析の加算・減算の影響

第3期の評価において加算・減算により、評価点の度数がどのように変化したかを図 3-1-3 に示す。全体として評価点 2 が減少し、評価点 3 が増加していることが読み取れる。

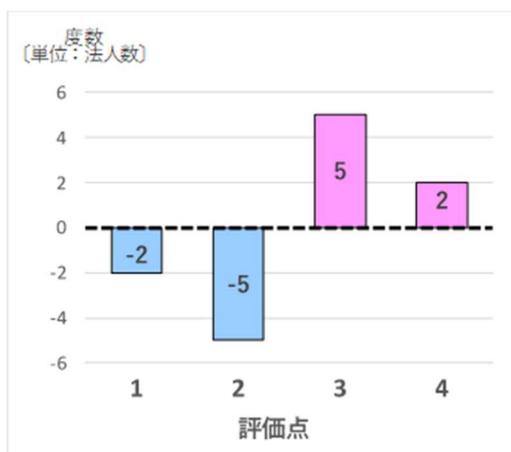


図 3-1-3 教育の達成状況評価点の度数の加算・減算後の増減

教育の達成状況評価点の、中期目標期間別の度数を図 3-1-4 に示す。第 1 期及び第 2 期ではほぼ評価点 2 に集中していたが、その傾向は第 3 期では和らぎ、特に評価点 1 と評価点 3 が増加していることが読み取れる。

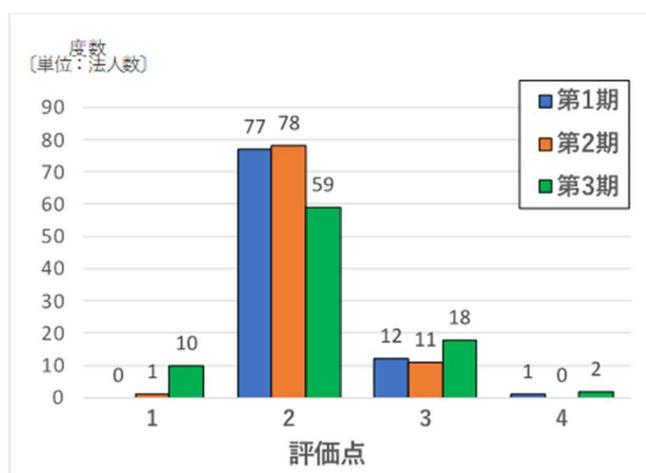


図 3-1-4 教育の達成状況評価点の中期目標期間別の度数

研究の達成状況評価結果への現況分析結果の加算・減算の影響について、第2期からの評価点の変化を図 3-1-5 に示す。教育の達成状況評価とは異なり、加算・減算前も第3期に評価点が向上しており、加算・減算後にはその傾向がさらに強くなっていることが読み取れる。

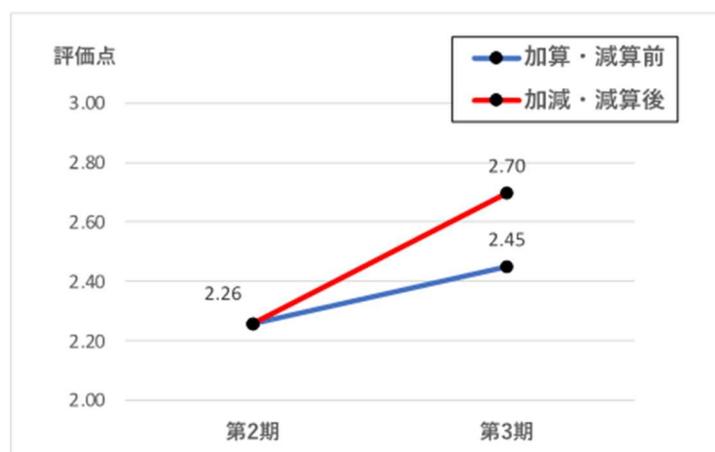


図 3-1-5 研究の達成状況評価の結果への現況分析の加算・減算の影響

第3期の評価において加算・減算により、評価点の度数がどのように変化したかについて図 3-1-6 に示す。全体として評価点4のみが15法人と大幅に増加していることが読み取れる。これも教育の評価点とは異なる傾向である。

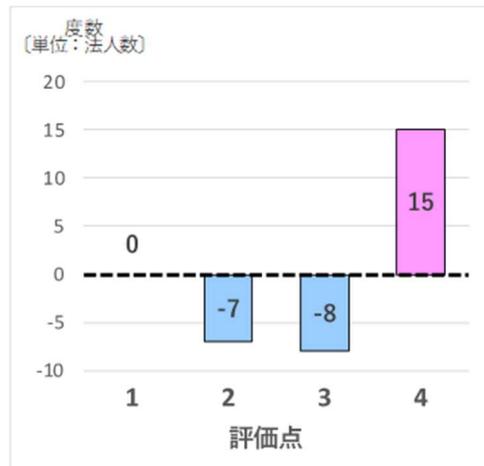


図 3-1-6 研究の達成状況評価点の度数の加算・減算後の増減

研究の達成状況評価点の、中期目標期間別の度数を図 3-1-7 に示す。第 1 期では評価点 2 に集中しており、第 2 期でもその傾向は強まっていたが、第 3 期では和らぎ、特に評価点 3 と評価点 4 が増加していることが読み取れる。一方で、評価点 1 も 6 法人と増加している。

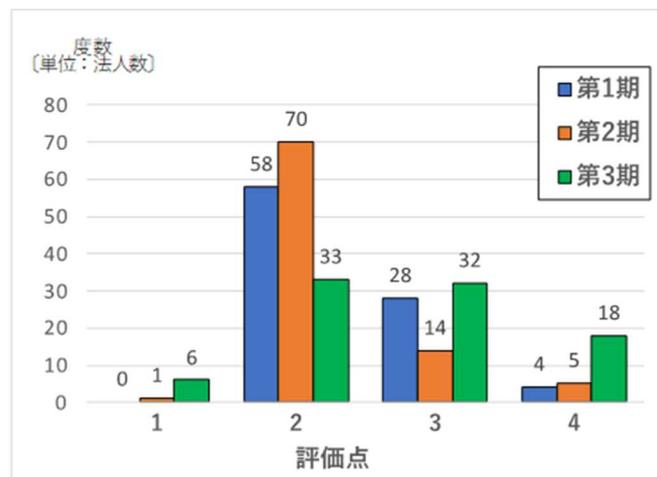


図 3-1-7 研究の達成状況評価点の中期目標期間別の度数

以上のデータから読み取れる傾向をまとめると、以下の 2 つが結論として言える。

1. 法人全体の平均値を見ると、教育の達成状況評価ではあまり変化がない一方で、研究の達成状況評価に関しては評価点が向上している傾向が見られた。
2. 法人全体の評価点のばらつきは教育・研究ともに大きくなり、評価点 2 に集中する傾向は和らいだ。

この結果を、評価方法の改善との関連で考察する。第 3 期に中期計画を 3 段階判定、中期目標の小項目を 5 段階判定とし、法人ごとに具体的な中期目標が設定されている小項目の評価を重視した仕組みとしたことは、現況分析結果の加算・減算前でも全体の評価点のばら

つきを第1期及び第2期と比べて一定程度大きくしたが、加算・減算によってさらにその傾向は強まった。

研究の達成状況評価において、第3期に評価点が向上しており、加算・減算によってさらにその傾向は強まった。これは今回分析において用いた評価点の定義による可能性もある。表3-1-2に示すように第3期では、「遅れている」と「重大な改善事項」に該当する判定はなかったことから、4段階の判定基準の評価点1として、第1期及び第2期の「不十分」と第3期の「おおむね順調」を同等として処理したが、これらの文言の意味的な違いは明確である。それゆえ、第3期では評価点1が増加したとも言えるだろう。第3期の大項目判定は小項目を積み上げて算出したため、評価者が大項目を直接判定する仕組みとはなっていなかったものの、達成状況の判定を示す記述が第1期及び第2期と同じ、「不十分」の表現であったら、評価を躊躇したであろうと想像できる。一方で、第3期の研究の達成状況の評価点の全体の平均値が加算・減算前も向上していることは、前項で考察したように、評価方法の変更、法人の研究力の向上、評価者の判断基準の変化など複数の要因が仮定され、本分析のみでは解釈が難しい。

また、本分析の結果は法人の達成状況評価結果を基にしたものであり、各法人が設定した目標の内容に依存するものであることにも留意が必要である。

以上のように、第3期において、達成状況の違いがより明瞭に判別できるようになり、評価方法の改善によって意図した結果が導かれたと言える。評価結果の活用や社会への説明責任を考える上で、評価結果の段階判定が数値化されて用いられる可能性を想定すると、段階の数の設定とその表現をどのようにするかは重要な問題である。判定を示す記述とその段階の数については、今後も第3期の変更による検証結果も含めて、引き続き慎重に検討していく必要がある。

2. 学部・研究科等の現況分析に関する分析

2.1 学系間での判定結果の差異

教育及び研究の現況分析結果（教育 865 学部・研究科等； 研究 591 学部・研究科等）に対して、分析項目（「活動の状況」、「成果の状況」）による違い及び学系（教育 10 学系，研究 11 学系）の違いによって判定結果がどのくらい異なっているかの検討を行った。

ここでは、教育及び研究の現況分析で同一の手法により統計的な分析を行った。データは全て機構のウェブサイトにて公表されており⁵、それらを基に各法人の学部・研究科等の評価結果から、判定を示す記述を基に点数化を行った。現況分析は4段階判定となっており、点数化は、「特筆すべき高い質にある」を4、「高い質にある」を3、「相応の質にある」を2、「質の向上が求められる」を1とした。

以上のように点数化したデータに対し、二元配置分散分析（混合計画）を適用した。被験者⁶内要因として分析項目（活動・成果）の2水準、被験者間要因として学系（教育 10 学系，研究 11 学系）について分析を行い、有意差が見られた場合には多重比較を行った。

2.1.1 教育に関する現況分析結果

現況分析結果の学系別平均値を教育活動の状況、教育成果の状況の別に集計した結果を図3-2-1に示す。なお、エラーバーは標準誤差を示している（図3-2-2も同様）。

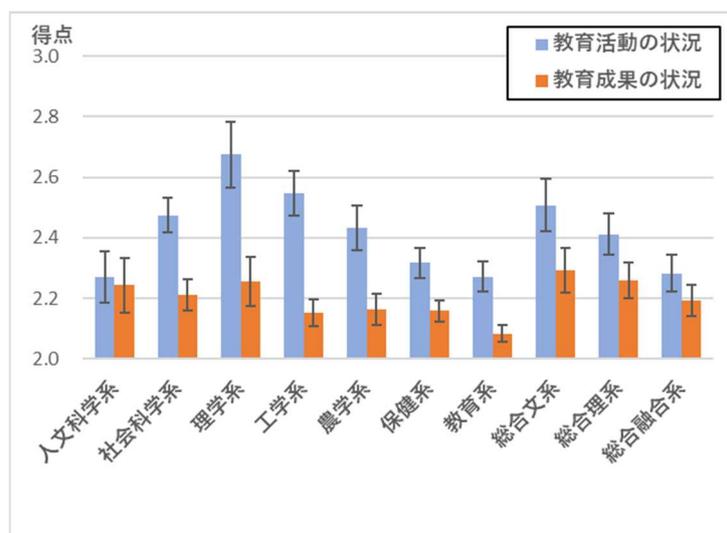


図3-2-1 教育に関する現況分析結果の学系別平均値

⁵ https://www.niad.ac.jp/evaluation/research_evaluation/kokuritikyoudou/hyoukakekka_R3/

⁶ ここでは学部・研究科等を被験者として扱って分析を行った。

分散分析の結果は

- ・ 学系の違いによる主効果が有意 $F(9, 855) = 2.89, p < .01, \eta_p^2 = .029$
- ・ 分析項目の違いによる主効果が有意 $F(1, 855) = 85.9, p < .01, \eta_p^2 = .091$
- ・ 交互作用は $F(9, 855) = 11.1, p = .012, \eta_p^2 = .024$

効果量を考慮し、学系の違い及び活動と成果の違いの主効果のみを解釈することとした。学系の違いについては、多重比較の結果、「教育系」-「理学系」($p = .013$)、「教育系」-「総合文系」($p = .048$)の間で有意差が見られた。

以上をまとめると、教育に関する現況分析の得点の平均値は、学系による差が見られ、「理学系」と「総合文系」が高く、「教育系」について低い傾向が見られた。しかし、その傾向はそれほど強くはないと解釈できる。また、分析項目については「教育活動の状況」の方が「教育成果の状況」よりも高い傾向が統計的に確認された。

2.1.2 研究に関する現況分析結果

現況分析の得点の学系別平均値を研究活動の状況、研究成果の状況の別に集計した結果を図 3-2-2 に示す。

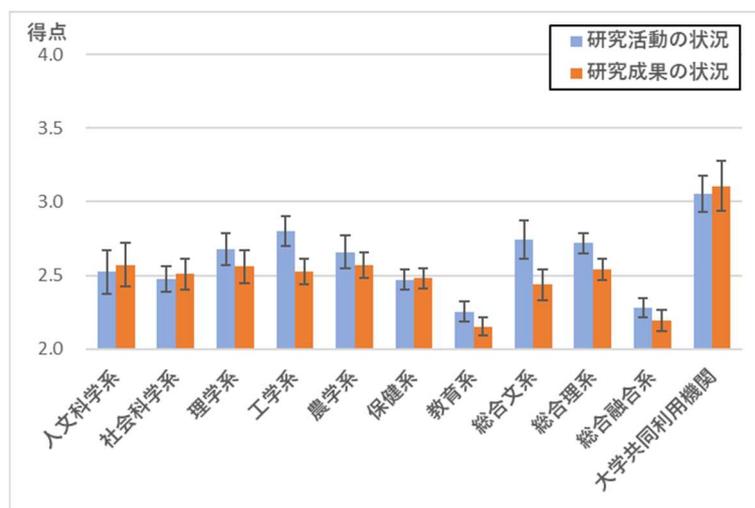


図 3-2-2 研究に関する現況分析結果の学系別平均値

分散分析の結果は

- ・ 学系の違いによる主効果が有意 $F(10, 580) = 4.61, p < .01, \eta_p^2 = .074$
- ・ 分析項目の違いによる主効果が $F(1, 580) = 5.79, p = .016, \eta_p^2 = .010$
- ・ 交互作用は $F(10, 580) = 2.33, p = .011, \eta_p^2 = .039$

効果量を考慮し、学系の違いの主効果のみを解釈することとした。学系の違いについては、

- ・ 「教育系」-「理学系」($p = .019$)、「教育系」-「工学系」($p < .01$)、「教育系」-「農学系」($p = .019$)、「教育系」-「綜合理系」($p < .01$)、「教育系」-「大学共同利用機

関」 ($p < .01$)

- ・ 「社会科学系」 - 「大学共同利用機関」 ($p < .01$)、 「保健系」 - 「大学共同利用機関」 ($p < .01$)、 「総合融合系」 - 「大学共同利用機関」 ($p < .01$)

で有意差がみられた。

以上をまとめると、研究に関する現況分析の得点の平均値は、学系の違いによる差が見られ、「大学共同利用機関」が高く、「教育系」について低い傾向が見られた。なお、分析項目については「研究活動の状況」と「研究成果の状況」の統計的な差は確認されなかった。

2.1.3 まとめ

以上 2 つの分散分析結果について考察する。「教育活動の状況」の方が「教育成果の状況」よりも高く評価されていたことは、「教育活動の状況」の方が「教育成果の状況」に比べ、教育の取組や活動などのプロセスの方が、エビデンスが明確で評価しやすいことが要因として考えられる。

学系ごとの違いについては、教育に関する現況分析の得点では、「理学系」と「総合文系」が高く、「教育系」について低い傾向が見られたが、それほど強くはなかった。一方で、研究に関する現況分析の得点では、「大学共同利用機関」が高い傾向が見られた。これは、大学共同利用機関の目的から考えると当然とも言える。一方で、「教育系」について低い傾向が見られた。これは、「研究成果の状況」では、その主たる根拠である研究業績説明書の評価において、科学研究費助成事業の小区分ごとに編成された専門部会において研究業績水準判定をした後に、現況分析部会の学系ごとにまとめることになっていることが影響している可能性がある。教育分野の研究の特色として、他分野にまたがった幅広い性質を持っていることや、加点要素となる特記事項を法人がアピールすることが他分野と比べて難しく、結果的に低い評価となった可能性もあるだろう。したがって、このような違いは、教育や研究の水準の直接的な差を表わすものではないことに留意されたい。

2.2 判定結果の第3期までの比較

前項では、第3期の判定結果について、学部等と研究科等の現況分析結果をまとめて分析したが、教育に関する現況分析結果については、学部等と研究科等で実施している教育内容が異なり、別個に分析することで詳細な傾向が明らかになる可能性がある。また、現況分析結果が第1期及び第2期からどのように変化したかについても検討することで新たな知見が得られる可能性がある。

以上を踏まえ、本項では教育組織の学部等と研究科等の教育の現況分析結果、研究組織の研究の現況分析結果について、分析項目ごとに、中期目標期間の違いによる評価結果の変化を分析した。教育の現況分析結果については第2期と第3期⁷、研究の現況分析結果については第1期、第2期及び第3期での差異を対象とした。前項に示したとおりデータは全て機構のウェブサイトにて公表されており、それらを基に各法人の学部・研究科等の評価結果から、判定を示す記述を基に点数化を行った。

なお、比較にあたり第1期と第2期では段階判定の区分とその考え方は同一の定義に基づいて行われていたが、第3期では水準の判定と質の向上度の判定を別々に行わず、水準の判定の中で質の向上の状況も含めて評価を行うように変更がなされた⁸。各期の判定基準に基づいた、点数化の比較表を表3-2-1に示す。

表3-2-1 各期の判定基準に基づいた分析項目の評価点への点数化

第1期および2期			第3期		
点数化	判定を示す記述	判断する考え方	点数化	判定を示す記述	判断する考え方
4	期待される水準を大きく上回る	取組や活動、成果の状況が非常に優れており、それぞれの学部・研究科等で想定する関係者の期待を大きく上回ると判断される場合	4	特筆すべき高い質にある	それぞれの学部・研究科等の目的に照らして、取組や活動、成果の状況が非常に優れていると判断される場合
3	期待される水準を上回る	取組や活動、成果の状況が優れており、それぞれの学部・研究科等で想定する関係者の期待を上回ると判断される場合	3	高い質にある	それぞれの学部・研究科等の目的に照らして、取組や活動、成果の状況が優れていると判断される場合
2	期待される水準にある	取組や活動、成果の状況は良好であり、それぞれの学部・研究科等で想定する関係者の期待に応えていると判断される場合	2	相応の質にある	それぞれの学部・研究科等の目的に照らして、取組や活動、成果の状況が相応であると判断される場合
1	期待される水準を下回る	取組や活動、成果の状況に問題があり、それぞれの学部・研究科等で想定する関係者の期待に応えられていないと判断される場合	1	質の向上が求められる	それぞれの学部・研究科等の目的に照らして、取組や活動、成果の状況が不十分であると判断される場合

分析においては、「総合文系」、「総合理系」、「総合融合系」、「大学共同利用機関」については、第1期及び第2期からの部会の構成の変更及び分析対象となる学部・研究科等

⁷ 教育に関する現況分析では第1期の分析項目の設定が異なり、比較できる形式のデータがないことにより、分析から除外した。

⁸ 第3期中期目標期間の教育研究の状況についての評価における「第2期からの主な変更点」, 独立行政法人大学改革支援・学位授与機構, 令和元年7月8日更新.

https://www.niad.ac.jp/media/006/201907/henkouten_201907.pdf

の数が少ないことを考慮して除外し、「人文科学系」、「社会科学系」、「理学系」、「工学系」、「農学系」、「保健系」、「教育系」の7学系を対象とした。

新設、改組、廃止等を除外し、中期目標期間の間で比較可能なデータを整理した結果、二元配置分散分析（対応あり）の分析対象とした教育研究組織は以下のとおりであった

- ・人文科学系（教育（学部等）20、教育（研究科等）17； 研究（学部・研究科等）21）
- ・社会科学系（教育（学部等）45、教育（研究科等）74； 研究（学部・研究科等）52）
- ・理学系（教育（学部等）26、教育（研究科等）15； 研究（学部・研究科等）34）
- ・工学系（教育（学部等）48、教育（研究科等）36； 研究（学部・研究科等）36）
- ・農学系（教育（学部等）38、教育（研究科等）27； 研究（学部・研究科等）31）
- ・保健系（教育（学部等）67、教育（研究科等）67； 研究（学部・研究科等）80）
- ・教育系（教育（学部等）45、教育（研究科等）65； 研究（学部・研究科等）47）

被験者内要因として分析項目（活動・成果）の2水準、中期目標期間の違いを教育では第2期、第3期の2水準、研究では第1期、第2期、第3期の3水準として設定し、分析を行った。

2.2.1 教育に関する現況分析結果（学部等）

学系別の現況分析評価点の平均値の変化を図3-2-3に示す。なお、このグラフは対応のあるデータに限定せず、該当する学系の各中期目標期間における全ての現況分析評価点を平均したものである。以下の図3-2-4、図3-2-5も同様である。

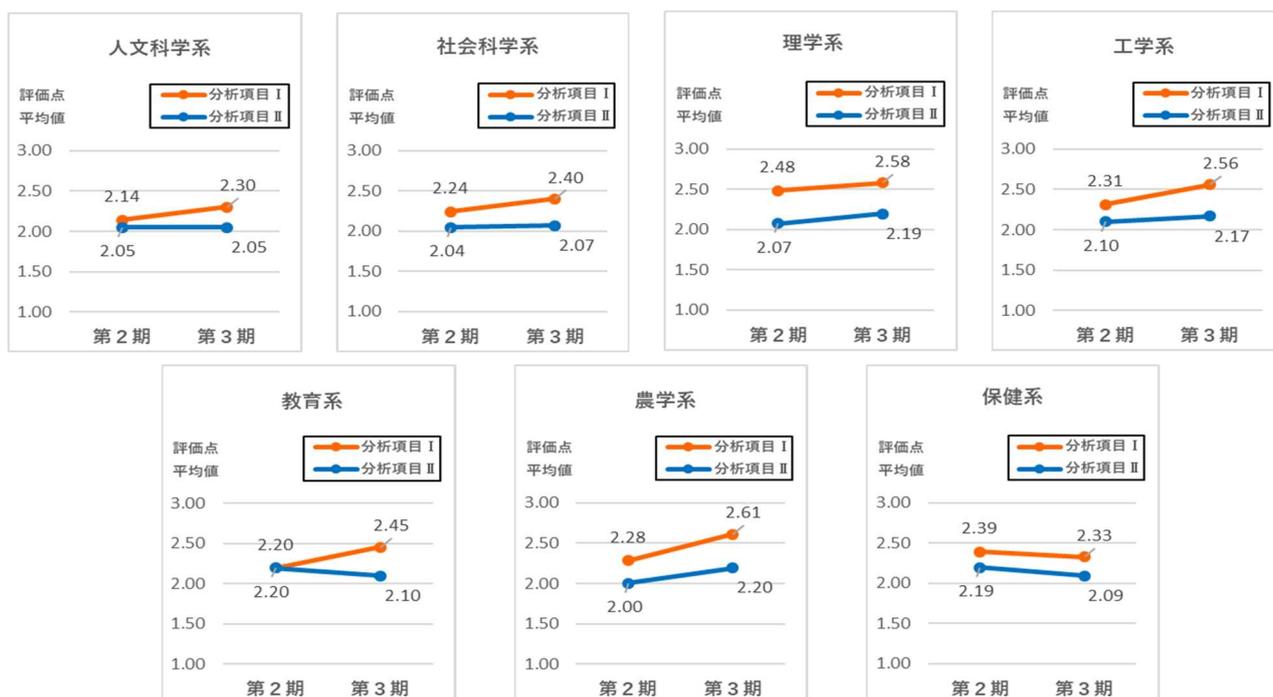


図3-2-3 学系別現況分析結果の変化（学部等 教育）

分散分析の結果は以下の通りであった。

・人文科学系

- 分析項目の違いによる主効果が有意 $F(1, 19) = 4.41, p = .049, \eta_p^2 = .188$
- 中期目標期間の違いによる主効果が $F(1, 19) = 1.00, p = .330, \eta_p^2 = .050$
- 交互作用は $F(1, 19) = 1.00, p = .330, \eta_p^2 = .050$

・社会科学系

- 分析項目の違いによる主効果が有意 $F(1, 44) = 26.6, p < .001, \eta_p^2 = .376$
- 中期目標期間の違いによる主効果が $F(1, 44) = 2.19, p = .146, \eta_p^2 = .047$
- 交互作用は $F(1, 44) = 2.67, p = .110, \eta_p^2 = .057$

・理学系

- 分析項目の違いによる主効果が有意 $F(1, 25) = 19.24, p < .001, \eta_p^2 = .435$
- 中期目標期間の違いによる主効果が $F(1, 25) = .632, p = .434, \eta_p^2 = .025$
- 交互作用は $F(1, 25) = .064, p = .802, \eta_p^2 = .003$

・工学系

- 分析項目の違いによる主効果が有意 $F(1, 47) = 24.24, p < .001, \eta_p^2 = .340$
- 中期目標期間の違いによる主効果が $F(1, 47) = 4.00, p = .051, \eta_p^2 = .078$
- 交互作用は $F(1, 47) = 2.89, p = .096, \eta_p^2 = .058$

・農学系

- 分析項目の違いによる主効果が有意 $F(1, 37) = 18.18, p < .001, \eta_p^2 = .329$
- 中期目標期間の違いによる主効果が有意 $F(1, 37) = 12.38, p < .01, \eta_p^2 = .251$
- 交互作用は $F(1, 37) = .886, p = .353, \eta_p^2 = .023$

・保健系

- 分析項目の違いによる主効果が有意 $F(1, 66) = 16.42, p < .001, \eta_p^2 = .199$
- 中期目標期間の違いによる主効果が $F(1, 66) = 2.08, p = .154, \eta_p^2 = .031$
- 交互作用は $F(1, 66) = 2.28, p = .635, \eta_p^2 = .003$

・教育系

- 分析項目の違いによる主効果が有意 $F(1, 44) = 4.18, p = .047, \eta_p^2 = .087$
- 中期目標期間の違いによる主効果が有意 $F(1, 44) = 4.18, p = .047, \eta_p^2 = .087$
- 交互作用は $F(1, 44) = 3.55, p = .066, \eta_p^2 = .075$

以上をまとめると、全ての学系で分析項目の違いによる影響が見られた。いずれも、「分析項目Ⅰ:教育活動の状況」のほうが、「分析項目Ⅱ:教育成果の状況」よりも高い傾向である。中期目標期間の評価点の平均値の違いは「農学系」「教育系」において見られ、いずれも第3期中期目標期間のほうが高くなっている。分析項目の別と中期目標期間の別の組み合わせから生じる影響（交互作用）は全ての学系において見られなかった。

2.2.2 教育に関する現況分析結果（研究科等）

学系別の現況分析評価点の平均値の変化を図 3-2-4 に示す。

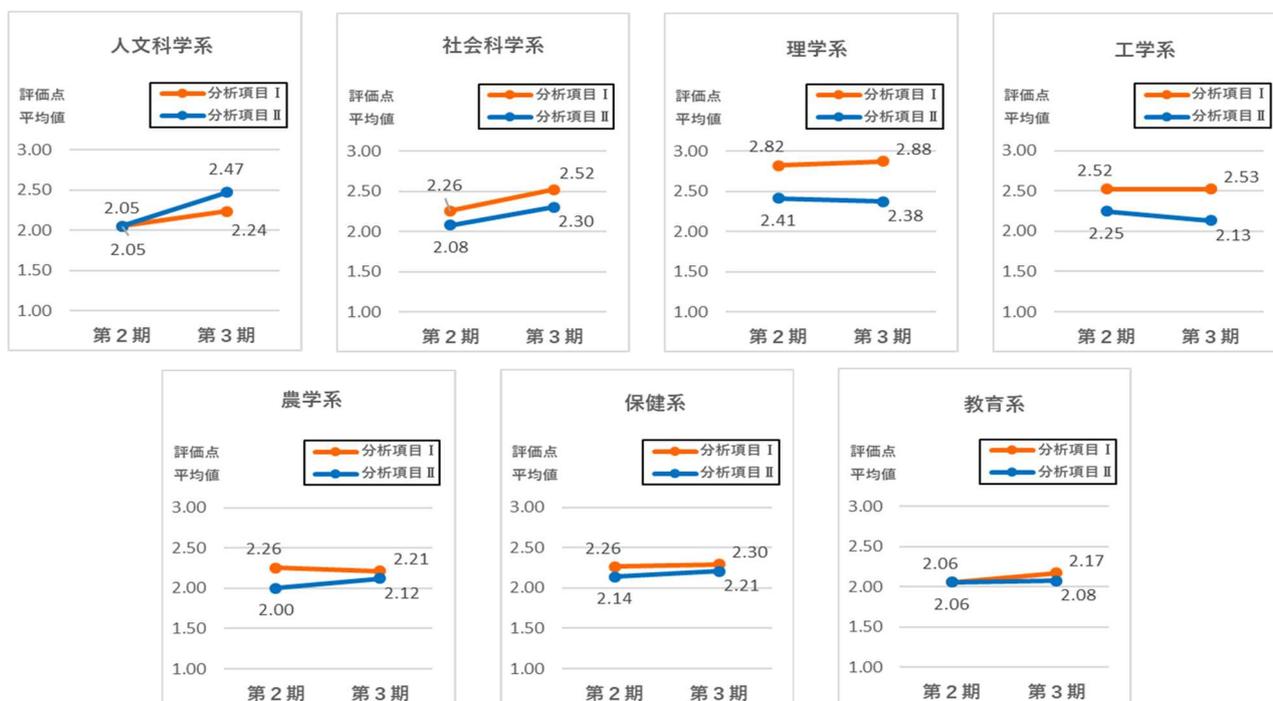


図 3-2-4 学系別現況分析結果の変化（研究科等 教育）

分散分析の結果は以下のとおりであった。

- ・ 人文科学系
 - 分析項目の違いによる主効果が $F(1, 16) = 1.66, p = .216, \eta_p^2 = .094$
 - 中期目標期間の違いによる主効果が有意 $F(1, 16) = 7.77, p = .013, \eta_p^2 = .327$
 - 交互作用は $F(1, 16) = 1.66, p = .216, \eta_p^2 = .094$
- ・ 社会科学系
 - 分析項目の違いによる主効果が有意 $F(1, 65) = 13.74, p < .001, \eta_p^2 = .175$
 - 中期目標期間の違いによる主効果が有意 $F(1, 65) = 8.69, p < .01, \eta_p^2 = .118$
 - 交互作用は $F(1, 65) = .116, p = .734, \eta_p^2 = .002$
- ・ 理学系
 - 分析項目の違いによる主効果が有意 $F(1, 14) = 26.39, p < .001, \eta_p^2 = .653$
 - 中期目標期間の違いによる主効果が $F(1, 14) = .00, p = 1.00, \eta_p^2 = .00$
 - 交互作用は $F(1, 14) = .483, p = .499, \eta_p^2 = .033$
- ・ 工学系
 - 分析項目の違いによる主効果が有意 $F(1, 35) = 21.05, p < .001, \eta_p^2 = .376$
 - 中期目標期間の違いによる主効果が $F(1, 35) = .103, p = .751, \eta_p^2 = .003$
 - 交互作用は $F(1, 35) = .722, p = .401, \eta_p^2 = .020$

・農学系

- 分析項目の違いによる主効果が有意 $F(1, 26) = 6.19, p = .020, \eta_p^2 = .190$
- 中期目標期間の違いによる主効果が $F(1, 26) = .325, p = .574, \eta_p^2 = .012$
- 交互作用は $F(1, 26) = 2.36, p = .136, \eta_p^2 = .083$

・保健系

- 分析項目の違いによる主効果が $F(1, 66) = 3.38, p = .070, \eta_p^2 = .049$
- 中期目標期間の違いによる主効果が $F(1, 66) = 1.07, p = .305, \eta_p^2 = .016$
- 交互作用は $F(1, 66) = .099, p = .754, \eta_p^2 = .001$

・教育系

- 分析項目の違いによる主効果が $F(1, 64) = 1.29, p = .260, \eta_p^2 = .020$
- 中期目標期間の違いによる主効果が有意 $F(1, 64) = 5.45, p = .022, \eta_p^2 = .079$
- 交互作用は $F(1, 64) = 1.51, p = .223, \eta_p^2 = .023$

以上をまとめると、「社会科学系」、「理学系」、「工学系」、「農学系」で分析項目の違いによる影響が見られた。いずれも、「分析項目Ⅰ:教育活動の状況」のほうが、「分析項目Ⅱ:教育成果の状況」よりも高い傾向である。また、中期目標期間期間の評価点の平均値の違いは「人文科学系」「社会科学系」「教育系」において見られ、いずれも第3期中期目標期間のほうが高くなっている。分析項目と中期目標期間の組み合わせから生じる影響（交互作用）は全ての学系において見られなかった。

2.2.3 研究に関する現況分析結果

学系別の現況分析評価点の平均値の変化を図 3-2-5 に示す。

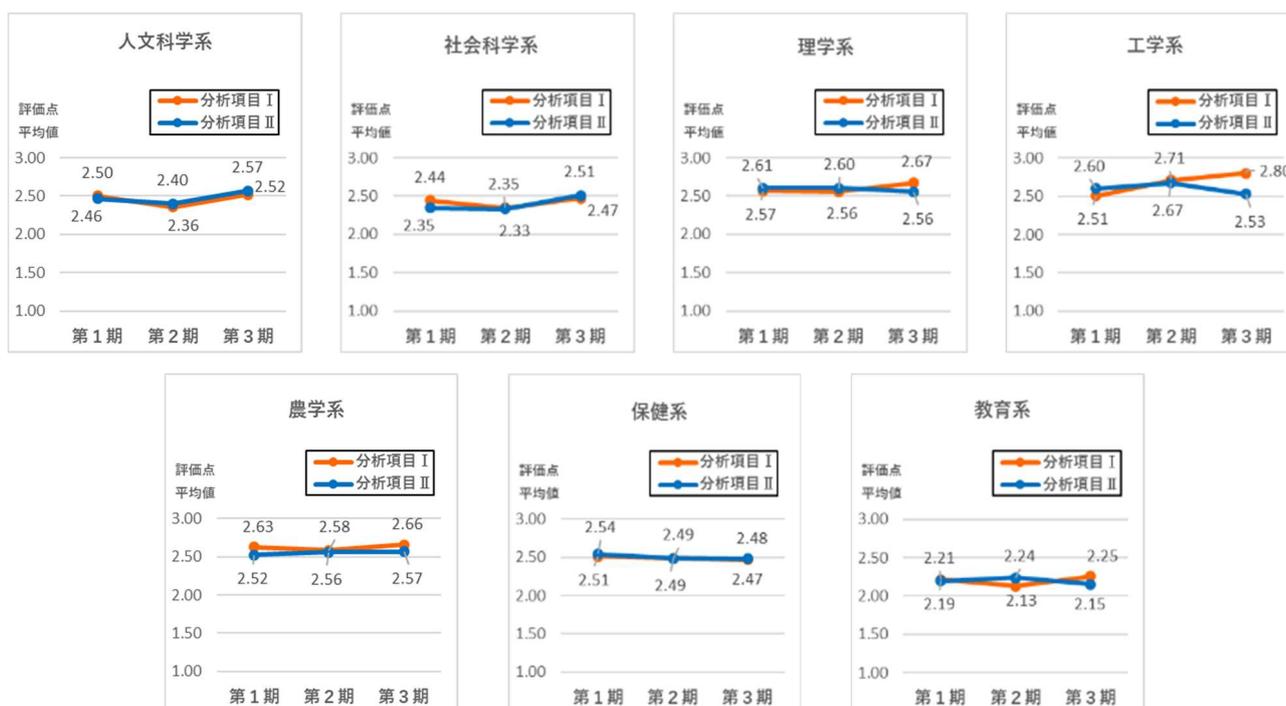


図 3-2-5 学系別現況分析結果の変化 (研究)

分散分析の結果は以下のとおりであった⁹。

・人文科学系

- 分析項目の違いによる主効果が $F(1, 20) = .137, p = .715, \eta_p^2 = .007$
- 中期目標期間の違いによる主効果が $F(2, 40) = .870, p = .427, \eta_p^2 = .042$
- 交互作用は $F(2, 40) = .426, p = .656, \eta_p^2 = .021$

・社会科学系

- 分析項目の違いによる主効果が $F(1, 51) = 1.04, p = .312, \eta_p^2 = .020$
- 中期目標期間の違いによる主効果が $F(2, 102) = 1.65, p = .197, \eta_p^2 = .031$
- 交互作用は $F(2, 102) = 1.93, p = .150, \eta_p^2 = .036$

・理学系

- 分析項目の違いによる主効果が $F(1, 33) = .195, p = .661, \eta_p^2 = .006$
- 中期目標期間の違いによる主効果が $F(2, 56) = .142, p = .835, \eta_p^2 = .004$
- 交互作用は $F(2, 66) = .555, p = .577, \eta_p^2 = .017$

・工学系

- 分析項目の違いによる主効果が $F(1, 35) = 2.25, p = .143, \eta_p^2 = .060$
- 中期目標期間の違いによる主効果が $F(2, 70) = .131, p = .877, \eta_p^2 = .004$
- 交互作用が有意 $F(2, 55) = 7.97, p < .01, \eta_p^2 = .185$

・農学系

- 分析項目の違いによる主効果が $F(1, 30) = .573, p = .455, \eta_p^2 = .019$
- 中期目標期間の違いによる主効果が $F(2, 60) = .108, p = .898, \eta_p^2 = .004$
- 交互作用は $F(2, 49) = .411, p = .624, \eta_p^2 = .014$

・保健系

- 分析項目の違いによる主効果が $F(1, 79) = .073, p = .787, \eta_p^2 = .001$
- 中期目標期間の違いによる主効果が $F(1.8, 142) = .441, p = .624, \eta_p^2 = .006$
- 交互作用は $F(1.8, 143) = .195, p = .801, \eta_p^2 = .002$

・教育系

- 分析項目の違いによる主効果が $F(1, 46) = .109, p = .743, \eta_p^2 = .002$
- 中期目標期間の違いによる主効果が $F(2, 92) = .025, p = .975, \eta_p^2 = .001$
- 交互作用は $F(2, 92) = 2.47, p = .090, \eta_p^2 = .051$

以上のように、「工学系」における交互作用のみ有意であった。単純主効果の検定を行ったところ、以下のとおりであった。

- ・ 分析項目 I において、中期目標期間の違いによる主効果が有意 $F(2, 70) = 3.06, p = .032, \eta_p^2 = .093$ 。多重比較の結果、第1期と第3期の間で有意差が見られ、第3期のほうが高くなっている ($p < .05$)。

⁹ 中期目標期間の効果と交互作用については、球面性の仮定が保持されない場合には、Greenhouse-Geisserの方法による自由度の補正を行った。

- ・ 分析項目 II において、中期目標期間の違いによる主効果は有意差なし $F(2, 55) = 3.23$, $p = .059$, $\eta_p^2 = .084$ 。
 - ・ 第 1 期中期目標期間において、分析項目の違いによる主効果は有意差なし $F(1, 64) = 1.82$, $p = .182$, $\eta_p^2 = .028$ 。
 - ・ 第 2 期中期目標期間において、分析項目の違いによる主効果は有意差なし $F(1, 54) = .197$, $p = .659$, $\eta_p^2 = .004$ 。
 - ・ 第 3 期中期目標期間において、分析項目の違いによる主効果が有意であり、分析項目 I の方が分析項目 II よりも高い傾向が見られた $F(1, 54) = 7.147$, $p = .010$, $\eta_p^2 = .117$ 。
- 以上から、「工学系」の第 3 期中期目標期間の分析項目 I 「研究活動の状況」について、相対的に高くなっていると解釈できる。

2.2.4 まとめ

以上 3 つの分散分析結果を基に考察を行う。

教育に関する現況分析の評価点においては、活動の方が成果よりも高い傾向であった。特に学部等でその傾向は強かった。これは、前項の第 3 期における学系間での現況分析得点の差異のみを対象とした検証と同様の結果であり、教育の取組や活動などのプロセスの方が、エビデンスが明確という考察との関連が考えられるだろう。

また、中期目標期間の評価点の平均値の違いについては、学部等の場合、「農学系」「教育学系」において第 3 期が高くなっており、一方で、研究科等の場合、「人文科学系」「社会科学系」「教育学系」において第 3 期が高くなっている。これは、第 3 期に関しては質の向上度も含めた評価であり、特記事項という加点要素が増えた評価の構造となったことその他、前項の達成状況報告書の分析で考察したのと同様に、評価方法、法人の教育の質の向上、評価者の判断基準の変化、など複数の要因が仮定され、本データの分析結果のみでは結論は付けられない。

第 3 期の「工学系」の研究に関する現況分析の分析では「研究活動の状況」が「研究成果の状況」よりも高い状況が見られた。これについても同様に複数の要因が仮定されるが、前項の図 3-2-2 を参照すると、今回分析対象としなかった「総合文系」、「総合理系」、「大学共同利用機関」においては、「工学系」と同様あるいはそれ以上の評価点であることが読み取れる。それゆえ、今回分析対象とした 7 学系内では「工学系」の「研究活動の状況」について高く見られたものの、全体としてみると学系による偏りが問題となるほどの現況分析結果とは言えないだろう。

このような違いは、教育や研究の水準の直接的な差を表すものとはいえないが、一方で、本項で実施したような学系ごとの分析をすることで、例えば、特定の学系に特徴的な数字が出てきた場合に、それを当該学系の判断として尊重すべきなのか、それとも調整を検討すべきものなのかの議論に際し、他の学系の分析結果を提供することによって、ピア・レビューの適切性を支援できる可能性が考えられる。

3. 研究業績水準判定に関する分析

本節では、研究業績水準判定に伴う、評定者間信頼性について統計的な分析を行った。複数の評価者が同一の研究業績を評価した際に、どの程度一致したか統計的に検証を行った。

最初に、検証の前提となる研究業績水準判定がどのように行われたかを、「評価作業マニュアル¹⁰」から引用する。

表 3-3-1 研究業績水準判定組織の作業および水準判断の基準（「評価作業マニュアル」より引用）

<p>●研究業績の水準判定組織の作業</p> <ul style="list-style-type: none">研究業績水準判定組織は、国立大学法人等から提出された「研究業績説明書」を分析します。分析は書面調査で行い、各研究業績の水準を判定し、現況分析部会及び達成状況判定会議へ提出します。研究業績水準判定組織は、専門委員によって構成します。評価を実施するために、研究分野別に専門部会を編成します。専門部会の研究分野の区分については、平成 31 年度(2019 年度)科学研究費助成事業 の分類における中区分を用います。研究業績水準判定組織が行う評価の実施体制は、各専門部会に「小区分」を定め、その「小区分」ごとに評価者を配置します。具体的な作業としては、各国立大学法人等より提出された「研究業績説明書」について、評価者 2 名が、研究テーマごとにそれぞれ水準を判定します。 <p>●研究業績の水準判断の基準</p> <p>判断は、1 学術的意義、2 社会、経済、文化的意義の視点で行ってください。</p> <p>国立大学法人等が、「学術的意義」「社会、経済、文化的意義」の双方を有する研究業績であると判断し、記載している場合は、双方の視点でそれぞれ判断してください。</p> <p>1 学術的意義での判断基準</p> <ul style="list-style-type: none">SS: 当該分野において、卓越した水準にあるS: 当該分野において、優秀な水準にあるS 未満: 以下の A、B、C に該当する A:当該分野において、良好な水準にある B:当該分野において、相応の水準にある(標準的な研究業績) C:上記の段階に達していない <p>2 社会、経済、文化的意義での判断基準</p> <ul style="list-style-type: none">SS: 社会、経済、文化への貢献が卓越しているS: 社会、経済、文化への貢献が優秀であるS 未満: 以下の A、B、C に該当する A:社会、経済、文化への貢献が良好である B:社会、経済、文化への貢献が相応である(標準的な研究業績) C:上記の段階に達していない

以上のように行われた研究業績水準判定において、「小区分」ごとの 2 名の評価者の「SS」、「S」、「S 未満」の判定があまりも一致していない場合や、一方の評価者が厳しすぎるまたは甘すぎる判定を行った場合には、判定結果の信頼性に疑問が持たれる。よって、2 名の評価者の判定の一致度はどの程度であったかを、統計的に分析した。

¹⁰ 評価作業マニュアル 国立大学法人及び大学共同利用機関法人の第 3 期中期目標期間の教育研究の状況についての評価（2020 年度実施：4 年目終了時評価），独立行政法人大学改革支援・学位授与機構，2018 年 6 月(2019 年 6 月改訂)， https://www.niad.ac.jp/media/006/201910/sagyoumanual_201910.pdf

3.1 方法

評定者間信頼性を現す指標として、一次の重み付き κ 係数を用いた。 κ 係数は、多くの応用がなされており、詳細は他の文献を参照されたいが¹¹、判定データに対してどのように適用したか、以下に簡潔に説明する。

2名の評価者の判定についてクロス表を作成すると、表 3-3-2 のように記述できる。 κ 係数とは、評定者の一致度を測る指標であり、一致度から偶然に一致する確率を考慮して除外したものである。今回のデータにおいては、順序尺度として「SS」、「S」、「S未満」として3つあるカテゴリにおいて、一方の評価者が「SS」で他方の評価者が「S」の判定だった場合（1段階の違い）と、一方の評価者が「SS」で他方の評価者が「S未満」の場合（2段階の違い）では、その不一致の程度が異なることから重み付き κ 係数¹²を用いることとした。 κ 係数は-1 から 1 までの値を取り、数値が高いほど判定の一致度が高くなることを意味している。 κ 係数の判断の目安の一つとして Landis and Koch (1977)による判断基準を以降では参考とすることとした¹³（表 3-3-3）。

表 3-3-2 研究業績水準判定における
2名の評価者のクロス集計表イメージ

		評価者 2		
		SS	S	S未満
評価者 1	SS	一致	ちょっと違う	全然違う
	S	ちょっと違う	一致	ちょっと違う
	S未満	全然違う	ちょっと違う	一致

表 3-3-3 κ 係数の判断の目安
(Landis & Koch, 1977)

κ 係数	一致の強さ
$\kappa \leq 0.0$	一致していない
$0.0 < \kappa \leq 0.2$	わずかに一致
$0.2 < \kappa \leq 0.4$	おおむね一致
$0.4 < \kappa \leq 0.6$	中程度に一致
$0.6 < \kappa \leq 0.8$	かなり一致
$0.8 < \kappa \leq 1.0$	ほとんど一致

研究業績水準判定では、それぞれの「小区分」につき2名の評価者が評価を行うことを基本としたが、業績数が多い一部の「小区分」では3名以上の評価者を配置する場合があった。その場合には、それぞれの評価者の組み合わせごとに κ 係数を算出し、分析においても個別の組み合わせとして扱った。科学研究費助成事業審査区分表（平成30年度）に拠れば、「小区分」の数は全てで306であったが、この処理により研究業績水準判定の組み合わせの数は361となる。また、 κ 係数を算出するためには、観測データの数が20-30程度必要なこ

¹¹ Kundel H.L., & Polansky M. (2003). Measurement of observer agreement. *Radiology*, 228, 303-308. などを参照のこと。

¹² 本分析では、判定の水準数が3と少ないため、判定の乖離が大きいときの不一致度をあまり重視しない1次の重み付き κ 係数を用いた。

¹³ Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.

とが経験的にわかっており (Hoel, 1976)¹⁴、本分析では 20 以上の研究業績が判定された組み合わせを分析対象 (以下、「小区分」という。) とした。

以上の手続きに基づくと、「学術的意義」で 248 小区分、「社会・経済・文化的意義」で 94 小区分であった。

3.2 小区分ごとに算出した κ 係数の全体の傾向

小区分ごとに算出した κ 係数を、表 3-3-4 の区分で集計したヒストグラムを図 3-3-1 に示す。階級の境界値は「超 - 以下」と設定した。

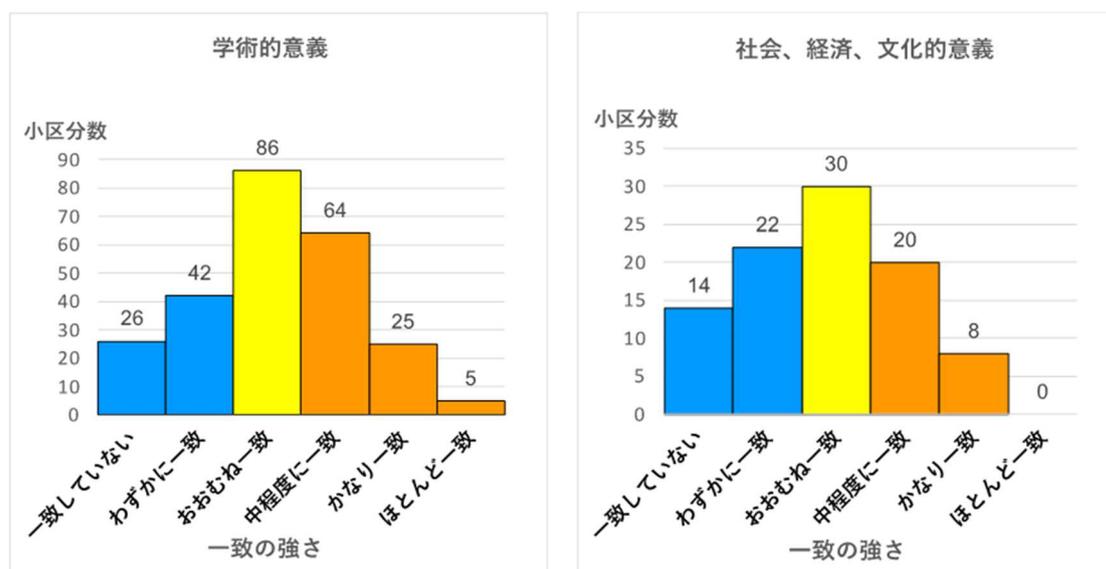


図 3-3-1 小区分における評価者間の一致の程度

「学術的意義」では、248 の小区分の内、「おおむね一致」が 86 (34.7%) と最も多かった。「中程度に一致」は 64 (25.8%)、「かなり一致」は 25 (10.1%)、「ほとんど一致」は 5 (2.0%) で、これにより一致度の高い「中程度に一致」以上は合計が 94 (38.0%) となった。一方、「わずかに一致」は 42 (16.9%)、「一致していない」は 26 (10.5%) で、一致度の低い「わずかに一致」以下は合計が 68 (27.4%) であった。

「社会、経済、文化的意義」では、94 の小区分の内、「おおむね一致」が 30 (31.9%) と最も多かった。「中程度に一致」は 20 (21.3%)、「かなり一致」は 8 (8.5%)、「ほとんど一致」は 0 (0.0%) で、一致度の高い「中程度に一致」以上は合計が 28 (29.8%) であった。一方、「わずかに一致」は 22 (23.4%)、「一致していない」は 14 (14.9%) で、「わずかに一致」以下は合計が 36 (38.3%) であった。

なお、 κ 係数の平均値は「学術的意義」で 0.34、「社会、経済、文化的意義」で 0.26 であり、標準偏差は「学術的意義」は 0.23、「社会、経済、文化的意義」は 0.24 であった。

¹⁴ Hoel, P.G. (1976). Elementary Statistics 4th Edition, Los Angeles, CA: Wiley

以上のように、「学術的意義」、「社会、経済、文化的意義」のいずれにおいても、「おおむね一致 ($0.2 < \kappa \leq 0.4$)」が最も多かった。

3.3 大区分ごとの κ 係数の傾向

ここでは、分野によって2名の評価者の判定の一致度に差があるか検討するため、科学研究費助成事業の大区分ごとに分析を行う。結果を解釈する上で、大区分にはどのような分野が含まれるかを理解する必要があるため、大区分と中区分について、科学研究費助成事業審査区分表（平成30年度）¹⁵より抜粋して整理したものを表3-3-4に示す。

以下に、大区分ごとの κ 係数の平均値の比較をするが、前項で述べたように、 κ 係数を算出するための基準を一つの研究業績水準判定の組み合わせで判定された業績数が20あることとしたため、大区分ごとに小区分の件数が異なった。その割合とともに件数を示す（図3-3-2）。これを見ると、「学術的意義」の大区分Kで3件、「社会、経済、文化的意義」の大区分Bで4件、Eで3件、Gで3件、Hで1件、Kで2件と小区分数が少なかったため、以下ではそれらを除外して分析を行った。

大区分ごとに該当する小区分の κ 係数の平均値を示す（図3-3-3）。「学術的意義」においては、大区分HとBにおいて κ 係数の平均値が0.44と「中程度に一致」していると解釈できる。これらの分野は主として薬学と理学に相当しており、大区分内での「小区分」の研究分野としての分類が、他の分野と比べて明確になっているが故に、2名の評価者の判定も一致する傾向が高く見られた可能性がある。一方で、比較的 κ 係数の平均値が低かった大区分J,D,Aについては、学際・融合的な傾向が強く、「小区分」内でも研究分野が学際的であるが故に、2名の評価者の判定の一致が低くなる傾向が見られた可能性がある。また、大区分Aについては人文・社会科学系という分野の特性も影響している可能性がある。

「社会、経済、文化的意義」においては、「学術的意義」と比べると、一番 κ 係数が高かった大区分Jで0.36、低かった大区分Aで0.23と、それほど全体としての違いが見られなかった。大区分Aが低かったことは「学術的意義」と同様に、人文・社会科学系という分野による影響の可能性もあるが、今回の分析においては、平均値を算出して比較した大区分に限られ、平均値の差も他の大区分と比べてもそれほど大きくないことから、本データのみでは判断はできない。

¹⁵ https://www.jsps.go.jp/j-grantsinaid/03_keikaku/data/h30/h30_bepyo2-1.pdf

表 3-3-4 研究業績水準判定における大区分と中区分

大区分	中区分	大区分	中区分
A	1 思想、芸術およびその関連分野	E	32 物理化学、機能物性化学およびその関連分野
	2 文学、言語学およびその関連分野		33 有機化学およびその関連分野
	3 歴史学、考古学、博物館学およびその関連分野		34 無機・錯体化学、分析化学およびその関連分野
	4 地理学、文化人類学、民俗学およびその関連分野		35 高分子、有機材料およびその関連分野
	5 法学およびその関連分野		36 無機材料化学、エネルギー関連化学およびその関連分野
	6 政治学およびその関連分野		37 生体分子化学およびその関連分野
	7 経済学、経営学およびその関連分野		F
	8 社会学およびその関連分野	39 生産環境農学およびその関連分野	
	9 教育学およびその関連分野	40 森林園科学、水圏応用科学およびその関連分野	
	10 心理学およびその関連分野	41 社会経済農学、農業工学およびその関連分野	
	42 獣医学、畜産学およびその関連分野		
B	11 代数学、幾何学およびその関連分野	G	43 分子レベルから細胞レベルの生物学およびその関連分野
	12 解析学、応用数学およびその関連分野		44 細胞レベルから個体レベルの生物学およびその関連分野
	13 物性物理学およびその関連分野		45 個体レベルから集団レベルの生物学と人類学およびその関連分野
	14 プラズマ学およびその関連分野		46 神経科学およびその関連分野
	15 素粒子、原子核、宇宙物理学およびその関連分野		H
	16 天文学およびその関連分野	48 生体の構造と機能およびその関連分野	
	17 地球惑星科学およびその関連分野		49 病理病態学、感染・免疫学およびその関連分野
C	18 材料力学、生産工学、設計工学およびその関連分野	I	50 腫瘍学およびその関連分野
	19 流体力学、熱工学およびその関連分野		51 ブレインサイエンスおよびその関連分野
	20 機械力学、ロボティクスおよびその関連分野		52 内科学一般およびその関連分野
	21 電気電子工学およびその関連分野		53 器官システム内科学およびその関連分野
	22 土木工学およびその関連分野		54 生体情報内科学およびその関連分野
	23 建築学およびその関連分野		55 恒常性維持器官の外科学およびその関連分野
	24 航空宇宙工学、船舶海洋工学およびその関連分野		56 生体機能および感覚に関する外科学およびその関連分野
	25 社会システム工学、安全工学、防災工学およびその関連分野		57 口腔科学およびその関連分野
	58 社会医学、看護学およびその関連分野		
D	26 材料工学およびその関連分野		59 スポーツ科学、体育、健康科学およびその関連分野
	27 化学工学およびその関連分野	90 人間工学およびその関連分野	
	28 ナノマイクロ科学およびその関連分野	J	60 情報科学、情報工学およびその関連分野
	29 応用物理物性およびその関連分野		61 人間情報学およびその関連分野
	30 応用理工学およびその関連分野		62 応用情報学およびその関連分野
	31 原子力工学、地球資源工学、エネルギー学およびその関連分野	K	63 環境解析評価およびその関連分野
90 人間工学およびその関連分野	64 環境保全対策およびその関連分野		

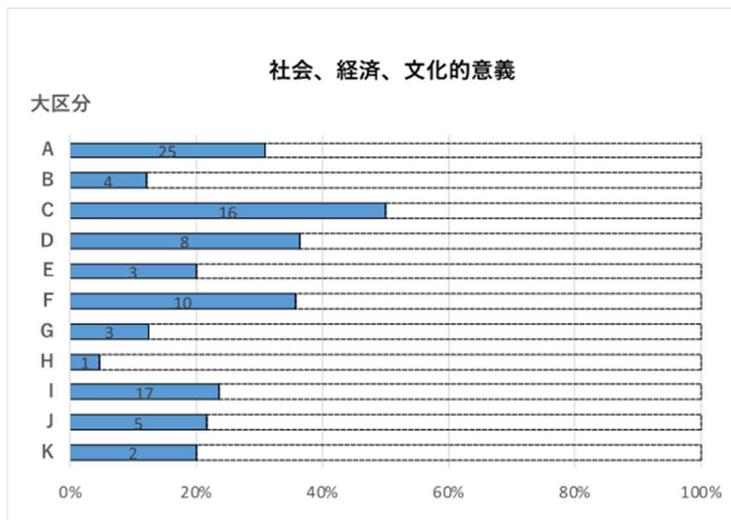
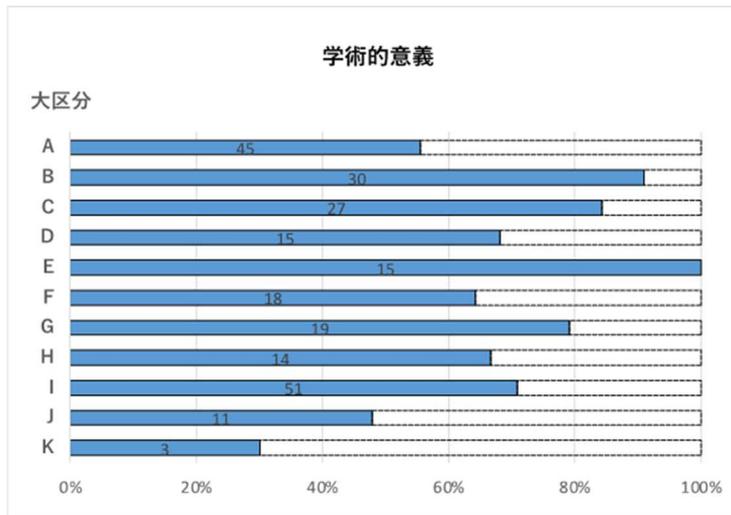


図 3-3-2 各大区分における、 κ 係数の算出を行った小区分数の全研究業績水準判定の組み合わせ数に占める割合と件数

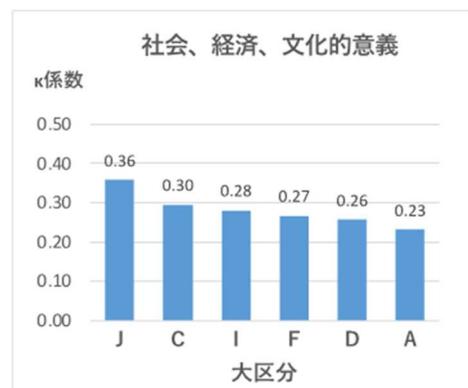
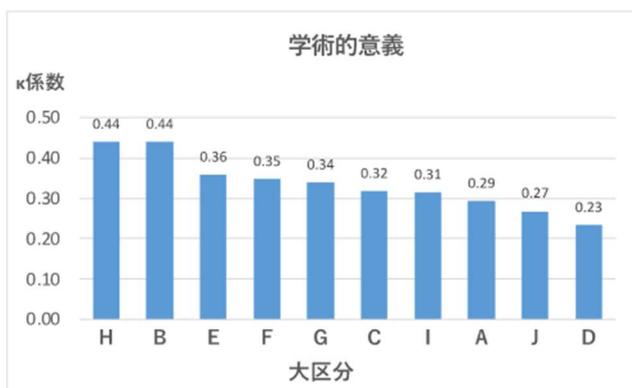


図 3-3-3 各大区分における κ 係数の平均値

3.4 まとめ

本節では、 κ 係数を算出することで、研究業績水準判定において2名の評価者の判定の一致度について検討を行った。その結果、

- ・ 「学術的意義」、「社会、経済、文化的意義」のいずれの評価においても、「おおむね一致 ($0.2 < \kappa \leq 0.4$)」が最も多かった。その一方で、「わずかに一致 ($0.0 < \kappa \leq 0.2$)」、「一致していない ($\kappa \leq 0.0$)」も一定数見られた。
- ・ 大区分ごとに該当する小区分の κ 係数の平均値を算出したものを比較した結果、「学術的意義」の評価において、学際・融合的な傾向が高いと、2名の評価者の判定の一致度が低くなる可能性が示唆された。

一方で、仮に2名の評価者の判断が一致しない場合でも、必ずしもその判定が否定されるものではない。例えば、評価者の選出の際に同じ「小区分」内で同じキーワードに代表される分野に偏らないように、バランスをとっていることを反映した結果である可能性があるだろう。そのような場合においては、評価者間の判定において一致の割合が高くなく、同一の研究業績に対する判定結果が相反しているとしても、それぞれの判定結果には一貫性が見られ、一定の信頼できる観点からの判定が行われている可能性も考えられる。

本節で実施したような小区分ごとの評価者間の判定の一致度の分析をすることで、例えば、特定の小区分で低い一致度が見られた場合に、前節で示した現況分析結果の学系間での差異の検討と同じく、ピア・レビューの判断として尊重すべきなのか、あるいは調整を必要とするべきものなのか、事後的な検証においてだけでなく、評価実施段階においても検討が可能となる可能性を有している。これらの分析手法の提案と本分析の結果は、より適切なピア・レビューを実施していく上での基礎的なデータとなると考えられる。

4. 定量的な指標と評価結果の関係

今回の評価では定量的な指標を、現況調査表作成時の「基本的な記載事項」の根拠となる資料・データの一部として活用した¹⁶。これらの大部を占めるデータ分析集は法人から毎年度収集するデータをもとに、機構において整理し、評価者に提供した。評価者に評価の根拠として用いる基礎的なデータを提供したことは、法人と評価者のエビデンスの共通化や評価に係る負担の軽減等を意図したものであった。

これらの定量的な指標は、その重要度が高いものから「重要指標」、「分析指標」、「参考指標」の3つに区分し、評価者にはそれぞれに異なる活用を求めている。ここでは、これらの指標と評価結果の関係を見るために、「重要指標」の中から「標準修業年限×1.5年内卒業（修了）率」と「入学定員充足率」を例にとり、教育に関する現況分析結果の「分析項目Ⅰ 教育活動の状況」及び「分析項目Ⅱ 教育成果の状況」との関係について分析した。

4.1 方法

「標準修業年限×1.5年内卒業（修了）率」と「入学定員充足率」については、今回の教育研究評価の対象期間である、平成28年度～令和元年度の平均値を算出した。「分析項目Ⅰ 教育活動の状況」及び「分析項目Ⅱ 教育成果の状況」の数値化は、表3-2-1の右表に沿って行い、本章2の分析で用いたものと同様に行った。なお、「保健系」のデータに関しては、「医学科」、「看護学科」等の学科単位で算出されており、教育組織との対応付けができないため、分析から除外した。また、「分析項目Ⅱ 教育成果の状況」の「人文科学系」、「総合理系」、「総合融合系」については、2判定に集中する傾向が非常に強く、相関係数の算出に適さないため除外した。

以上のように算出したデータに対し、学系別に、重要指標の平均値と分析項目の判定結果との間でのポリシリアル相関係数を算出した。ポリシリアル相関係数は、比尺度と順序尺度の間での相関係数として用いられる(Olsson et al., 1982¹⁷)。

¹⁶ 詳細については「現況調査表ガイドライン」を参照のこと。

https://www.niad.ac.jp/media/006/201907/genkyou_guidelines_201907.pdf

¹⁷ Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, 47(3), 337–347.

4.2 結果

4.2.1 「標準修業年限×1.5年内卒業（修了）率」と教育に関する現況分析結果との関係

「標準修業年限×1.5年内卒業（修了）率」と教育に関する現況分析結果の「分析項目Ⅱ 教育成果の状況」とのポリシリアル相関係数を示す¹⁸（表 3-4-1）。「理学系」、「工学系」、「教育系」において、中程度の正の相関関係が見られ、検定結果も有意であった。これらの散布図を図 3-4-1 に示す。グラフからわかるとおり、「教育系」の関係では少数の4判定に影響されることから解釈に注意が必要である。「理学系」と「工学系」においても、2判定に集中している傾向はあるものの、中程度の正の相関関係があると解釈も可能である。

表 3-4-1 「標準修業年限×1.5年内卒業（修了）率」と教育に関する現況分析結果
（分析項目Ⅱ 教育成果の状況）とのポリシリアル相関係数

学系	分析項目Ⅱ
社会科学系	0.224
理学系	0.555*
工学系	0.581**
農学系	0.268
教育系	0.717**
総合文系	0.544

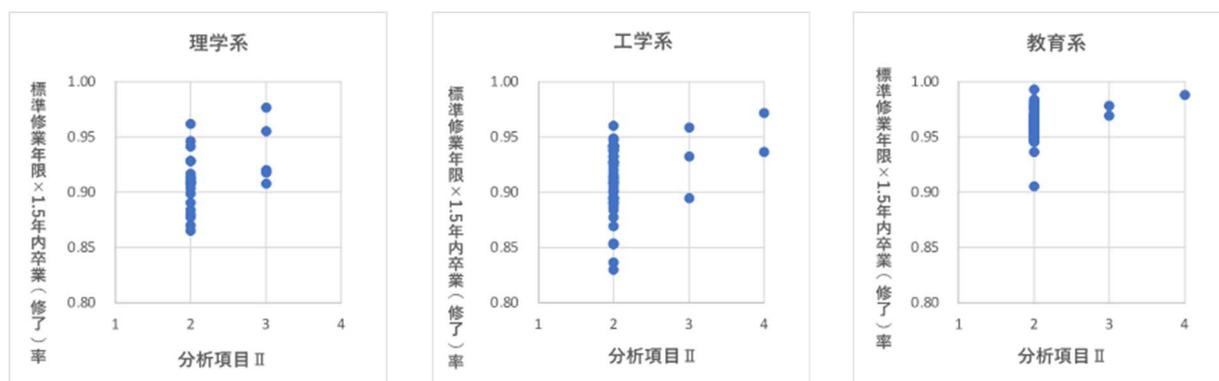


図 3-4-1 「標準修業年限×1.5年内卒業（修了）率」と教育に関する現況分析結果
（分析項目Ⅱ 教育成果の状況）との散布図（一部学系抜粋）

4.2.2 「入学定員充足率」と教育に関する現況分析結果との関係

「入学定員充足率」と教育に関する現況分析結果の「分析項目Ⅰ 教育活動の状況」とのポリシリアル相関係数を示す（表 3-4-2）。「教育系」に中程度の負の相関が見られ、「総合理系」との関係で強い正の相関関係が見られた。しかしながら、図 3-4-2 を見ると、「総合理系」においては、入学定員充足率が 0.72 である 2 判定の大学に影響を受けていること

¹⁸ 相関係数の有意性検定を行った結果、 $p < .05$ を*, $p < .01$ を**で示す。

がわかる。よって、「教育系」のみに負の相関関係があるとも解釈可能であるが、全体の学系の個数と相関の程度を考えると、全体としては「入学定員充足率」と「分析項目Ⅰ 教育活動の状況」との間には、あまり相関が見られなかったと解釈するべきであろう。

表 3-4-2 「入学定員充足率」と教育に関する現況分析結果
(分析項目Ⅱ 教育活動の状況)とのポリシリアル相関係数

学系	分析項目Ⅰ
人文学系	0.474
社会科学系	-0.015
理学系	-0.321
工学系	0.199
農学系	0.079
教育系	-0.453**
総合文系	0.142
総合理系	0.809**
総合融合系	0.320

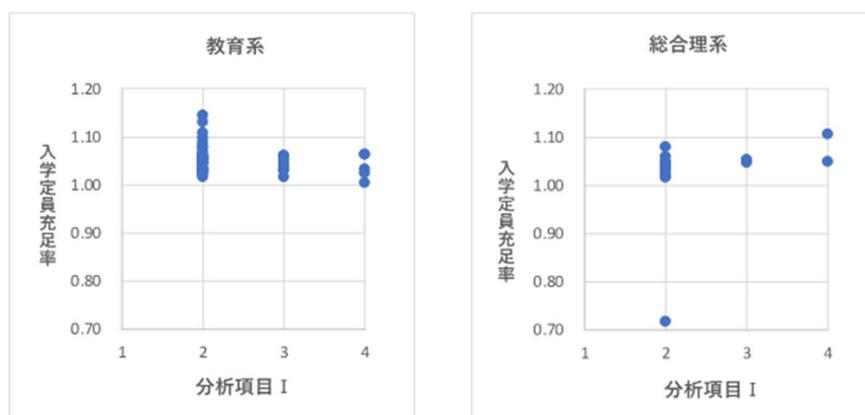


図 3-4-2 「入学定員充足率」と教育に関する現況分析結果
(分析項目Ⅱ 教育活動の状況)との散布図 (一部学系抜粋)

4.3 まとめ

「標準修業年限×1.5年内卒業(修了)率」と「分析項目Ⅱ 教育成果の状況」の間では、「理学系」と「工学系」において中程度の正の相関関係が見られた。「入学定員充足率」と「分析項目Ⅰ 教育活動の状況」との間では、相関はあまり見られなかった。

両者の違いをもたらした原因として、現況調査表ガイドラインに提示した必須記載項目が「分析項目Ⅱ 教育成果の状況」では少なかったことから、「標準修業年限×1.5年内卒業(修了)率」が重要な根拠として作用した可能性が考えられる。「分析項目Ⅱ 教育成果の状況」における必須記載項目は「1 卒業(修了)率、資格取得等」と「2 就職、進学」の2項目であった一方で、「分析項目Ⅰ 教育活動の状況」では「1 学位授与方針」、「2 教育課程方針」、「3 教育課程の編成、授業科目の内容」、「4 授業形態、学習指導法」、「5 履修指導、支援」、「6 成績評価」、「7 卒業(修了)判定」、「8 学生の受入」の8項目

が設定されており、全体の判定の中での指標の重みが違うことが、違いを生じさせた原因の1つとして考えられる。

また、相関が見られなかったことは、必ずしも指標が評価に用いられていないことを意味するわけではない。一つは、今回の教育研究評価で提案した重要指標であっても、単一の指標のみで、評価結果が説明できるわけではないことを反映した結果の可能性もある。現況分析の段階判定では「特記事項」による加点が重視されており、定量的な指標はそれを補完する位置づけとなっていることによる可能性もある。さらに、分析に用いたデータの特性として、分析項目の段階判定は2判定が中心であり、3判定や特に4判定は少数であったことにより、相関を見るという今回の分析の枠組みでは把握できなかった可能性もある。

効果的・効率的な評価の実施において、適切な指標設定は不可欠である。引き続き、指標の設定について精査を続けていく必要がある。

【調査実施体制】

土屋 俊	研究開発部 部長	(全体総括)
渋井 進	研究開発部 教授	(第Ⅱ章及び第Ⅲ章の検証・執筆)
赤川 裕美	調査研究補佐員	(第Ⅲ章に関するデータ分析補助)
浅井 美紀	調査研究補佐員	(第Ⅱ章に関するデータ分析補助)
評価事業部評価企画課	国立大学評価室	(アンケート集計、資料編纂)

【問い合わせ先】

独立行政法人大学改革支援・学位授与機構
評価事業部評価企画課 国立大学評価室
042-307-7907, 7908, 7913
houjin2@niad.ac.jp