

【原著】

AO 入試における多面的・総合的評価の基準担保を目指して

——集団討論のための評価者研修会およびセミナーを通じての考察——

菊池明泰，細川和彦，塚越久美子，碓山恵子，石田眞二（北海道科学大学）

2015 年に策定された高大接続改革実行プランをもとに，北海道科学大学では個別選抜のうち AO 入試においてルーブリックを用い多面的・総合的評価を基盤とした「新ガリレオ入試」を導入した。生徒一人ひとりの多様な能力を「公平」に評価する選抜で，3 回のセミナー（講義，集団討論，実験・実習など）を通じて生徒を評価する制度である。この入試制度の導入から今年で 3 年が経過したが，受講生増加に伴い評価者も増員させてきた。そこで評価者間の評価基準の担保が重要となり，その対策としてセミナー開始前に評価者を対象とした 3 回の研修会を実施した。その研修内容の有用性について実際の評価結果をもとに考察した。

1 はじめに

2014 年に中央教育審議会による答申が出されてから，現在にいたるまで高大接続改革実行プランの策定（2015），高大接続改革の実施方針等の策定について（2017）など，大学個別選抜に関わる多くの資料が提示されている。個別選抜に多面的・総合的評価の考え方を取り入れた入試制度導入の例として，2010 年度から講義と実験見学，レポート提出をもとにした選抜法を取り入れた東京農工大学農学部環境資源科学科の「ゼミナール入試」，アサーティブプログラムとよばれる個別面談や基礎学力・集団討論などによる選抜法を取り入れた追手門大学の「アサーティブ入試」などが挙げられる。これらの入試制度は，従来のペーパーテストや面接のみによる個別選抜とは異なり，受験する生徒の特性を，いろいろな課題をもとに評価する選抜法である。課題を総合的に評価するにあたり，どのような基準で評価し，さらに課題ごとの評価の重みをどのように割り振るかが重要となる。特に私立大学では，さまざまな入試区分が存在し，その違いを明確に受験生に提示することが必要となり，大学ごとに策定されている「アドミッション・ポリシー」との整合性も考慮しつつ入試制度を考える必要がある。

北海道科学大学（以下，本学）は，新しい個別選抜の方法の一つとして平成 28 年度より「新ガリレオ入試（AO 入試）」を導入した。これは，AO 試験前に講義を機軸に集団討論や実験・実習などの課題をセミナー形式で受講してもらい，それらの課題を総合的に評価し AO 入試のための出願許可を出したうえで入試を行う方法である。

2017 年，平成 30 年度入試として 3 回目を迎えたこの制度は高校生への認知度も高くなり，一定の評価を得たことで参加人数も初回と比べ大幅に増加した

（図 1）。

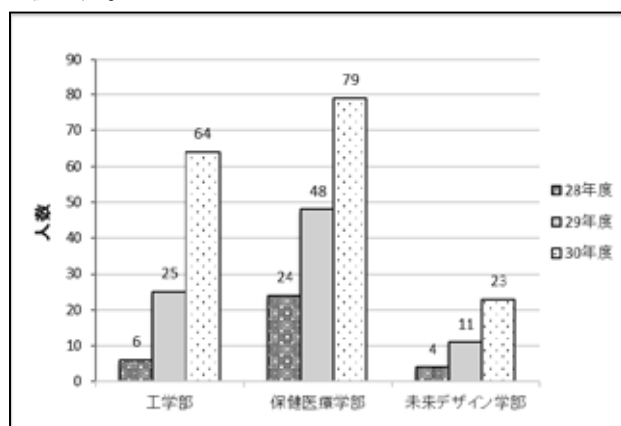


図 1 各年度別の受講者の推移（学部別）

初年度は全学で 34 名であったが，30 年度の受講数は約 5 倍の 166 名の生徒が受講し，それに伴い評価する側（以下，評価者）も人数を増員する必要性が出てきた。

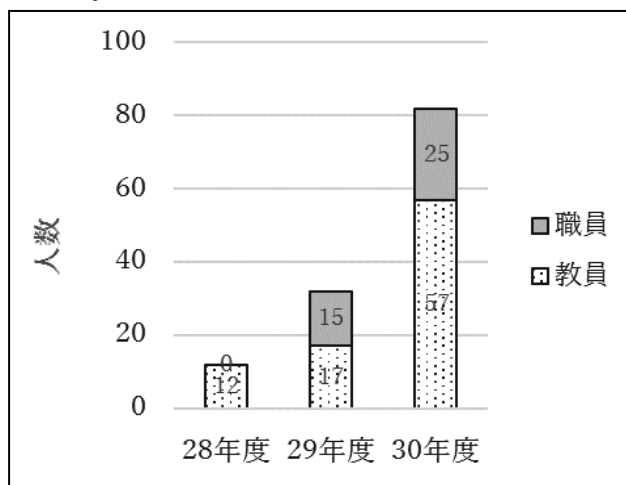


図 2 年度別評価者数の推移

ここで、問題となるのが評価者間の評価基準の担保である。集団討論や実験・実習では 30 年度合計で 80 名を超える評価者が必要となり（図 2），評価のばらつきが学内において懸念された。そこで，28 年度入試からセミナー前に評価者研修を年 3 回実施し，集団討論の映像を見ながら，実際の評価で使用するルーブリックにより模擬評価を行ったり，ルーブリックの改善点を挙げたり，見直しなどを進めている。

本研究では，集団討論の評価基準の担保が研修会によって実現可能か，30 年度入試の研修会の評価者ごとの結果と実際のセミナー時の結果をもとに比較検証したので報告する。

2 新ガリレオ入試（AO 入試）

2.1 入試の概要

本学で新たな入試制度として平成 28 年度入試よりスタートした「新ガリレオ入試（AO 入試）」は，物事を科学的な手法で検証していくという，本学の建学の精神の考え方をもとに，16 世紀に活躍したイタリアの科学者「ガリレオ・ガリレイ」の自ら実験を行い自分の理論を証明するという思想を現代に照らし合わせて命名した本学独自の入試制度である。

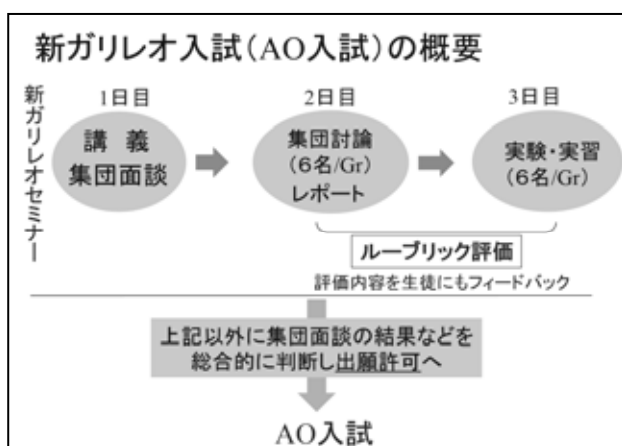


図 3 新ガリレオセミナーの概要

図 3 に示すように，3 つのセミナーと，その後に実施される AO 入試で構成されており，セミナー受講生（以下，受講生）は 3 回のセミナー全てに参加する必要がある。1 回目のセミナーは，「講義」「集団面接」「個別面接（評価には含まず）」，2 回目は「集団討論」と「レポート」，3 回目は「実験・実習」であり，それぞれ 1 日ずつ 3 日間実施する。このうち，「集団面接」と「個別面接」以外はすべてルーブリック評価を用いている。なお，最終日の「実験・実習」では，受講生それぞれが実験・実習での成

果をプレゼンする場面も設定している。受講生は各自が希望する学科の所属学部（工学部・保健医療学部・未来デザイン学部）ごとにセミナーを受講する。最終的にすべての課題と「集団面接」を点数化し，基準をクリアした受験生に AO 入試への出願が認められる。

2.2 班編成と課題の評価について

2 回目セミナーの「集団討論」と 3 回目セミナーの「実験・実習」はどちらも同じメンバーで班を編成した。1 班 6 名で男女の割合や希望学科などの偏りが生じないように考慮した。2 つの課題を同じ班員にすることで，3 回目の実験・実習時に必要な班員間の協働性を保つことができる。図 4 の集団討論では，班ごと



図 4 工学部の集団討論の様子

に付箋や各自でまとめたノートをもとに意見を出し合い 3 回目の実験・実習に向けた内容をまとめている。

集団討論，実験・実習の評価については，導入初年度より，各班 2 名の評価者（教員 1 名・職員 1 名）があらかじめ作成したルーブリック評価表を用いている。そして 29 年度からは受講者の増加とともに集団討論の評価に，30 年度からは実験・実習の評価にも本学の職員が評価者として加わっている。

ルーブリックは点数化をすることが難しい集団討論や，レポートなどを客観的に評価できるもので，縦横の表内に評価観点と評価レベルを簡潔な文章に記載したものである。セミナーでは，このルーブリックを用い 4 段階で評価したが，文章では理解が難しい部分もあり，言葉の捉え方によって評価点が変わってしまうことも想定された。評価基準が評価者に左右されることが個別選抜では重要であり，特に集団討論など点数化が難しいものでは，公平に評価しなければなら

ない。そのため本学ではセミナー開始前に評価者を対象とした研修会を複数回実施することにした。

2.3 多面的・総合的評価の担保に向けて

2.3.1 評価者のための研修会

評価者のための研修	
・ 1回目:新ガリレオ入試の概要とルーブリック評価について ⇒アクティブラーニング形式によるルーブリック評価表の作成	
・ 2回目:模擬集団討論による評価の実際 ⇒本学学生による集団討論とその評価 -学部別での実施 -討論のテーマ・討論時間・教室の大きさなども実際のセミナーと同じ条件で設定	
・ 3回目:視覚教材を用いた評価 ⇒2回目の集団討論時に撮影した動画を用いた評価 ⇒評価者からの意見交換とルーブリック評価表の見直し	

図5 研修の概要

各研修会の内容について図5に示す。研修会では主に集団討論についても評価観点について研修を行った。1回目の研修では、簡単なルーブリック評価表を実際に評価者間で作成することで、評価表の見方や考え方などの理解を深めた。2回目の研修では実際に本学の学生にボランティアとして参加してもらい6つの班を作り集団討論を実施。評価者が本番で使用するルーブリック評価表を用いて評価を実施した。模擬のグループの構成についても、本番のセミナーを想定し活発なグループや比較的静かな学生などをランダムに配置する形で実施した。終了後には、評価した班ごと



図6 動画画面の抜粋（実際の画面では学生の表情がわかるようになっている）

に分かれ、評価の結果について意見交換も実施した。3回目の研修では2回目の集団討論をビデオ撮影した

ものを編集し、同一学生を複数で評価する方法で研修を実施した（図6）。この動画教材については学生の表情や周りに対するコメントなどがわかるように編集し、表情や身振りなどさまざまな視点から評価できるようにした。なお、2回目の研修時複数の評価者より「教室が狭く班どうしが近いと、発言している学生の声が聞こえづらい」、「評価者が動きづらい」などの指摘があり、実際の集団討論を行う教室は通常の人数と比べ倍の広さの教室を用いることにした。

その後、2回目と3回目での結果をもとに本番で使用するルーブリック評価表について、評価者から意見を集約し、文言の確認・修正を実施した。

2.3.2 ルーブリック評価表の改訂

図7に28年度（初回）と29年度のルーブリック評価表の一部抜粋したものを示す。1回目の反省点と研修会での意見をもとに、コミュニケーションの部分

平成28年度			
	目標	A	B
コミュニケーション力	他者の意見を共感し聴くとともに、意見を引き出すことができる	メンバー全員と活発に意見を交わし、議論を深めている	意見交換を活発に行い議論に参加している
判断力	課題解決のためにすべきことが正確に判断できる	討論のまとめ役となってメンバーをリードし、積極的に討論をまとめている	他のメンバーの意見を引き出し、討論を進めている
平成29年度			
	目標	A	B
コミュニケーション力（傾聴）	他者の意見を聴くことができる	他者の意見を引き出すことができる	他者の意見を共感を持って聴ける
コミュニケーション力（発言）	積極的に発言することができる	積極的に発言し議論を深めている	他者の意見と噛み合った発言をしている
判断力	意見構築のためにすべきことが正確に判断できている	討論のまとめ役となってメンバーをリードし、積極的に討論をまとめている	討論の流れを理解し、討論に貢献している

図7 集団討論のルーブリック評価表（抜粋）

を「傾聴」と「発言」の二つ細分化し、評価者の誰が見てもわかりやすい文言とした。またレポートや実験・実習に対する評価表についても見直しを図った。評価表変更時の共通の見解として、評価を判定する項目内の文章量をできる限り少なくすることを念頭に置いた。評価は、評価観点毎に到達目標を設定した上で、評価自体は4段階評価で各項目5点満点としている。さらに受講生に対しては、この表をもとにしたフィードバック表を新たに作成し、セミナー終了後に返却し、実際のセミナーで各自がどの程度到達できたかが分かるようにした。なお、30年度のセミナーでは、ルーブリック評価表の変更点はなく同じ評価表を用いた。

3 評価の一致性について

3.1 研修時での評価結果

3回目の研修会では動画を用いて同一の学生評価を

全評価者で実施し、ルーブリック評価表による評価結果の一致性について検証した。

検証方法については動画を一定時間視聴後、評価者ごとの評価が集計可能なアンサーパッドを用い、口頭で以下の 2 点について質問し、評価結果を集計した。

- ① 時間経過とともにどのようにグループ内で発言しているか【発言力】
- ② 他者の意見をどのように聞きとっているか【傾聴力】

質問と集計は、ディスカッションを加えながら複数回実施し、評価は実際と同じ A・B・C・D の 4 段階に分け、A から順に、とてもよい・よい・普通・問題あり、とした。

その結果を図 8 - ①, ②に示す。なおこの研修には、教員 57 名、職員 16 名の合計 73 名が参加した。

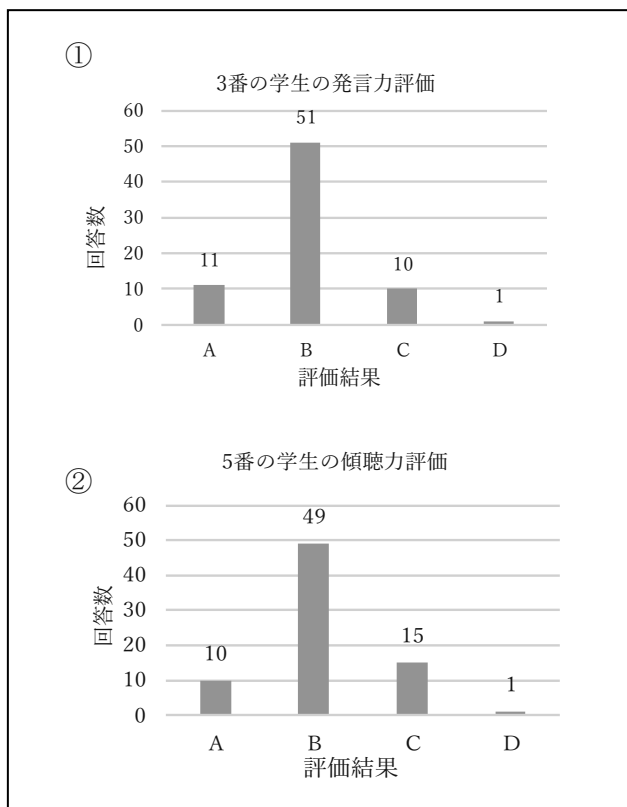


図 8 学生ごとの評価結果

①の発言力評価については、図 6 の 3 番の学生の評価結果だが、ほとんどの評価者が A「大変よい」もしくは B「よい」と評価し、評価者全体の 85%を占めた。他の学生と比べ発言回数自体は多くない学生ではあったが、発言内容が的確であったとの意見が多かった。

一方、②の傾聴力評価は、図 6 の 5 番の学生について評価し、結果は A および B を合計すると評価者

全体の 78%がよい評価としたが、一方で C の「普通」とした評価者は 20%であった。

3.2 セミナーでの評価検証結果

実際のセミナーにおいても集団討論における評価者間の評価の一致率について分析し、3 つの学部（工学部・保健医療学部・未来デザイン学部）および全学部で検証した。

セミナーでは、評価は 1 班 5-6 名の受講生に対し、2 名（1 名教員、1 名職員）で行っている。評価時間は 20 分の集団討論が 2 回で、評価項目は発言力・傾聴力のほか判断力・分析力の 4 項目について評価を行った。一致率の検討方法は、級内相関係数（表 1）および評価点差の絶対値の平均・標準偏差・最大値にて比較検討を行った。なお、級内相関係数の算出については JMP13 を用いた。表 1 および表 2 に結果を示す。

表 1 評価者間の級内相関係数

	評価者数	受講者数	級内相関係数
工学部	22	60	0.89
保健医療学部	26	77	0.80
未来デザイン学部	8	23	0.44
全学部	56	160	0.75

表 2 評価者間の点数差

	平均	標準偏差	最大値
工学部	1	1	5
保健医療学部	2	2	6
未来デザイン学部	3	3	10
全学部	2	2	10

級内相関係数は 1 に近いほど評価の一致率が高いが、全学部では 0.75 とある程度一致していた。学部別で一番高い一致率を示したのは工学部で、0.89 とかなり高い一致率を示した。一方、低い値を示したのは未来デザイン学部の 0.44 であった。

2 人の評価者間の点数差では工学部の平均が 1 と少なく、未来デザイン学部が 3 と高い値となり、最大では 10 点の差があった。

4 考察

多面的・総合的な評価を行う本学の「新ガリレオ入試」における評価基準の公平性を担保するために、評

価を担当する教職員 73 名を対象に研修会を実施した。

評価者の研修時での一貫性については、同一学生を多くの評価者で時間をかけて評価することで、発言内容や傾聴力を詳細にみる時間があり、結果的に評価に差が出なかったと思われる。また、今回評価の対象となった学生は、学内で集団討論を実際に行ったことがあるため、討論の仕方や他の班員の意見を聞く姿勢についても経験があった。それが評価者の相当数が良い評価を出した要因といえる。一方で傾聴力については、図 8 のように、「普通」とする評価者も一定数いた。ここで評価者からは「人の話を聞く」という行為自体を外から見たとき、「うなずきが多いだけで評価が高いといえるのか」、「聞いているだけで、その意見に対し返答しなかった場合には、どのように評価すべきなのか」など具体的な事例をあげた質問が出された。実際の評価では、「他の班員の意見をもとに、そこから自分の中で内容を咀嚼し、新たに意見を出し合えるスキルがあれば、傾聴力が高いと言えるのではないか」という意見も研修会の中で出された。

また研修時の集団討論では、消極的な学生の割合が多い班もあり「班内での相対的な発言力・傾聴力をもとに評価しても問題ないのか」、「班編成の方法によっては評価基準にも影響を及ぼすのではないか」との意見が出された。集団討論の評価では、評価者がすべての受講者を見て評価することは時間的に不可能なため、担当する班員のみでの相対的な評価になってしまう。そのため積極的な受講生が多い班と、消極的な受講生のほうが多い班では、班全体の集団討論の質が大きく変わる可能性もある。それを可能な限り防ぐため、今回の研修では、学科ごとに班を編成し積極的な班員がいるところと、消極的な班員が多い班をあえて編成し、研修を実施した。各班の雰囲気の違いによって評価に違いが出るかどうかを検証するためである。これを踏まえ、実際のセミナーでは、セミナー1日目に実施する個別面談での内容や男女の比率などを総合的に考慮し、できるだけ偏りが無いよう班編成を行った。

実際のセミナー評価者間の一致率については、学部別では工学部の一致率が一番高く、未来デザイン学部が一番低い値となった。工学部の一致率や評価者間の点数差が少ない要因としては、今年度の受講生が全般的に積極的だったため、評価自体が全体的にとってもよく、高い点数であったためと推測される。テーマが「ドローン」に関するもので、内容自体も集団討論の課題として発展性に富むものであったことも要因と思われる。保健医療学部の場合は、評価者のうち評価を全体的に「普通」とつける評価者と「よい」とつける

評価者が一定数いたため、一致率が若干低くとなったと推測される。

一方、未来デザイン学部では、一部の班で「大変よい」とつけた評価者と、「普通」「問題あり」とする評価者がいたため、結果一致率が低いものとなった。これは、未来デザイン学部が文理融合の学部であり、文系や理系、デザイン系など受講者の個性が多様であった点も背景にあったと推測できる。討論の途中でグループ討論が紛糾した班があり、その際、同じ受講生の対応について、リーダーシップや分析力・発言力があると高く評価した評価者がいた一方、協調性や傾聴力に欠けると評価した評価者がいたことで、結果的に評価にバラつきがでてしまった面がある。また、他学部と比較して受講生の絶対数が少なかったことも、相対評価がある程度極端になってしまった理由といえるかもしれない。

しかしながら、最終的には、研修時の評価一貫性がよかったこと、実際のセミナー時の全学における一致率も高い値であったことなどを踏まえると、研修会を実施することで集団討論の評価基準の担保は、ある程度達成できたといえるのではないかと考える。

5 まとめと今後の課題

今年度で 3 回目を迎えた本学の AO 入試「新ガリレオ入試」は、初年度 30 名程度からスタートし今年度は約 5 倍の 166 名の生徒が受講する入試となった。多面的・総合的な評価を入試に取り入れるため、集団討論や実験・実習を導入し、様々な形式の課題提示や評価方法を模索しながら、評価法ではルーブリック評価を用い、透明性のある公平な評価を目指す入試制度を目指してきた。今回のように 160 名を超える受講者が対象の場合、集団討論などの班編成の方法、他の班員の声が影響しない広さの部屋の確保、そして多人数評価者に対応した評価基準策定の重要性が、この 3 年間の入試と研修会を通じて確認できた。

本稿では、その中で評価者の増加に伴う評価者間の評価基準の担保のため、評価の一貫性を 3 回の研修と実際のセミナー結果で比較検証した。複数の評価者の結果が一致することが重要なのかという議論もあるが、複数の評価者が一定の基準で評価することは多面的・総合的に、そして公平に個別選抜する観点に立つと重要であると考え検証を行った。

今回の評価者研修と実際のセミナーでの評価結果の検証からは、評価については全体としてはかなり高い一致率を得ることができた。評価基準の担保には完全ではないが、ほぼ成功しているといえるのではないかと

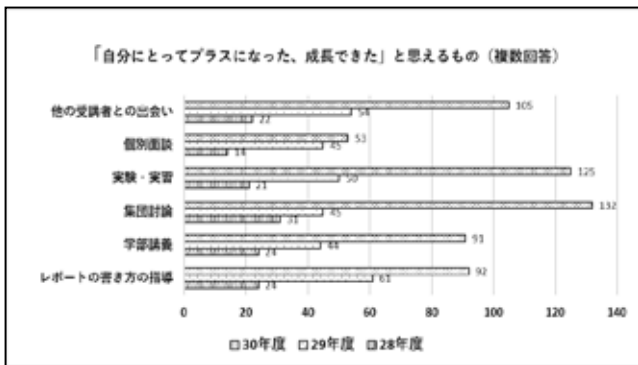


図9 受講者からの意見 (3年間のアンケート結果)

と考える。

しかし、一部の学部では一致率が低い値であったこと、さらに評価者人数が増加した場合、研修内容の質担保をどうするかなど、今後も検討項目があると思われる。日々の業務の中で多数の評価者の研修時間を確保する難しさがあがり、また3回のセミナーについても同様に時間と労力がかかる。こうした点から考えれば、今後は公平性や評価基準の一定の担保を維持しつつ、より効果的な入試制度自体を改めて模索することが求められといえるだろう。しかしながら、我々が3年間、セミナー受講生に対して行った同じ質問内容・無記名によるアンケート(図9)によれば、集団討論や実験・実習は自分自身の成長に非常にプラスになったと答えた受講生が多くいる。さらに、まだ具体的な検証には至っていないが、この新ガリレオ入試で入学した学生の入学後の意欲やその後の伸びは、他の入試制度よりも高いという手ごたえも各学科で感じられている。高大接続改革にも好影響を与えるものだと実感できるものがあるのは確かであり、今後も引き続き取り組むべき入試制度、評価方法であると考えている。一方で、セミナーで取り組むテーマの選定や提示の方法などについては、より受講生にわかりやすくするため工夫が必要であるとも感じた。また北海道という広範囲な地域性のため、遠方から3回のセミナーを札幌の本学に来校して受講するのは、受講生にとっては時間的にも金銭面でも負担が大きく、本学以外での実施や日程についても考慮すべきであると思われる。

最後に、今後も評価者研修会の内容を改善しながら、より公平な評価を目指し、個別選抜を実施していきたい。

謝辞

本報告にあたり、新ガリレオ入試に評価者として関わった北海道科学大学の教職員の方々、そして大学入

試広報センター主任の方々、川名氏をはじめ入試第一課の職員の方々、またセミナーに参加いただきアンケートに協力いただいた受講生の方々に深謝いたします。

参考文献

- 中央教育審議会答申(2014)「新しい時代にふさわしい高大接続の実現に向けた高等学校教育、大学教育、大学入学者選抜の一体的改革について～すべての若者が夢や目標を芽吹かせ、未来に花開かせるために～」
http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/_icsFiles/afieldfile/2015/01/14/1354191.pdf (2017年10月2日)
- 文部科学省高等教育局長 平成27年度大学入学選抜実施要項について(2014)
http://www.mext.go.jp/component/a_menu/education/detail/_icsFiles/afieldfile/2016/06/10/1282953_04.pdf (2017年10月2日)
- 文部科学省 高大接続改革実行プラン(2015)
http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo12/sonota/_icsFiles/afieldfile/2015/01/23/1354545.pdf (2017年10月2日)
- 東京農工大学 環境資源科学科 平成27年度～29年度ゼミナール入試実施内容
http://web.tuat.ac.jp/~enrs/ao/H30ao_naiyou.pdf (2018年1月3日)
- 追手門学院大学 2016年度アサーティブプログラム・アサーティブ入試 補助事業報告書
https://www.otemon.ac.jp/library/assertive/plan/pdf/2016_report.pdf (2018年1月3日)
- ダネルースティーブンス・アントニアーレビ(2014). 「大学教員のためのルーブリック評価入門(高等教育シリーズ)」玉川大学出版部
- 西岡加名恵, 石井英真, 田中耕治(2015). 「新しい教育評価入門」, 有斐閣
- 菊池明泰, 塚越久美子, 碓山恵子, 細川和彦(2017). ほかに「AO入試における多面的評価の導入—ルーブリック評価を用いた入試制度の構築—」『大学入試研究ジャーナル』, 27, 23-28.
- 菊池明泰(2017). 「AO入試『新ガリレオセミナー』」文部科学教育通信 407号 16-17.
- 菊池明泰(2017). 「AO入試『新ガリレオセミナー』のためのルーブリック評価」文部科学教育通信 409号 22-23.