

ルーブリックを用いた成績評価に関する報告 —教養学部「総合研究」の事例—

岸 浩介（教養学部言語文化学科）

1. はじめに

近年、大学の教育現場でも、組織的に、成績評価方法の一つとして「ルーブリック」を用いることが増えてきている。「ルーブリック」とは、たとえば、中央教育審議会（以下、中教審）（2012）の答申における「用語集」では「米国で開発された学修評価の基準の作成方法であり、評価水準である「尺度」と、尺度を満たした場合の「特徴の記述」で構成される」と説明されている¹。また、ステューブンス／レビ（2014: 2）では「ある課題について、できるようになってもらいたい特定の事柄を配置するための道具」および「ある課題をいくつかの構成要素に分け、その要素ごとに評価基準を満たすレベルについて詳細に説明したものであり、様々な課題の評価に使うことができる」と述べられている。言い換えれば、ルーブリックとは、ある学修活動や学修成果の評価などを行うさいに、その活動・成果に含まれるいくつかの指標の達成基準を明示し、その活動・成果全体の評価に利用するものであり、卒業論文などの成果を評価する際にも有用であると言えよう。本学教養学部でも、その必修科目である「総合研究（卒業課題）A」「総合研究（卒業課題）B」（いわゆる卒業論文または卒業研究）の評価の際に共通のルーブリックが学部全体で利用されている²。

後述の通り、この共通のルーブリックは、2019年度の試用の後、本格的運用が行われてきたが、教養学部では教員による実際の評価結果を可能な範囲で収集してきた。その評価結果の収集データのうち、2021年度の評価結果に基づき、筆者は2022年3月11日に東北学院大学泉キャンパスで開催された「2021年度 第2回 教養学部 FD研修会」において報告を行った。本論はその時の報告に基づいたものである。

本論の構成は以下の通りである。まず、第2節で、「総合研究」の概要とともに、これまで実施してきた「総合研究」のルーブリック評価データ収集についての目的と経緯を簡単に紹介

¹ 「用語集」は、中教審（2012）に含まれている。リンクは以下の通りである（2022年10月31日閲覧）。

https://www.mext.go.jp/component/b_menu/shingi/toushin/_icsFiles/fieldfile/2012/10/04/1325048_3.pdf

² 「総合研究（卒業課題）A」「総合研究（卒業課題）B」は、以降、単に「総合研究A」「総合研究B」または両者をまとめてタイトルにもあるように「総合研究」と略記する。

する。続く第3節では、2021年度の評価データ収集の過程を紹介し、学科および入試種別ごとのデータの概要とともに、学生が「総合研究」で身につけるべき知識・技能を全体的にどの程度身につけたと言えるのか、また、評価結果と最終的な「総合研究」の成績との相関関係について述べる。最後の第4節では、本論の結語とともに今後の課題を記す。

2. ルーブリック評価データ収集の背景・目的・経緯

2.1. 背景—教養学部の「総合研究」とは—

本節では、「総合研究」の概略を説明する。教養学部では、4年次必修の学部共通科目として「総合研究」(4単位)を設定している。この「総合研究」は、各学科における専門的知識・技能・思考法のみならず、教養学部の教育理念の一部である「学際性」ならびに「総合性」による「複合的視点」を活用しつつ「課題を発見し解決する能力」を所属学生に身につけさせることを目的とした科目である³。本論で扱う評価結果を収集した2021年度は、教養学部所属の全教員87名のうち78名が「総合研究」を担当しており、ほぼ全ての教員が関わっていると言える。担当教員は、複数の教員で構成される「チーム」に属し、協働的に所属学生の指導にあたる。「チーム」は、学科内の教員で構成されるものもあれば、数は多くないものの学科横断的に形成されるものもある。所属学生は、毎年度、所属チームにおいて原則として年2回の発表会を経て、所定の期日に論文や作品などの成果物を提出した後、口頭試問や成果発表会などを経て、『総合研究(卒業課題)論文要旨』という論文要旨集に要旨を掲載する。

2.2. 目的

次に、どのような目的のもとで、学部全体でルーブリックを使用することになったのかについて述べる。上述の通り、「総合研究」は、複数のチーム編成のもとで進行する。各チームでの教育内容ならびに成果物が満たすべき要件、評価基準・方法はそれぞれの専門分野での基準や慣例に基づいている。これらを「全ての側面で」「完全に」統一化・画一化することは不可能かつ非合理的ではあるものの、全教養学部生が4年間の学修成果を総合的に活用しながら履修する必修科目である以上、評価基準にはある程度の統一性が必要になるだろう。

加えて、ここ数年、大学教育でも学修成果を評価するための取り組みが全国的に展開されるようになってきている。その契機の一つとなっているのは、中教審(2012)の答申における「学士課程教育の質的転換」を促すための「学修成果の可視化」であり、ルーブリックがその方法

³ 東北学院大学教養学部(2017)の「学位授与の方針」(ディプロマ・ポリシー)ならびに「教育課程編成・実施の方針」(カリキュラム・ポリシー)を参照。

の一つとしてあげられた⁴。

このような状況を受け、教養学部でも「総合研究」の共通ルーブリックを策定・活用しようとする取り組みが2018年度に開始された。その取り組みはそれ以降も続き、教養学部は現在も、本学における中・長期的プロジェクトであるTG Grand Vision 150の中で、ルーブリックを活用した「総合研究」の学修成果の可視化を中期的目標の一つに掲げている。このような変革の促進は、大学をとりまく周囲の要請による部分が少なくないが、このような（一定程度の）統一的评价基準を設けることにより、学生が4年間の学修で身につけたと思われる知識・能力・技能をある程度客観的に測定できることには少なからぬ意義があることは言うまでもない。

2.3. 経緯

次に、本節では、「総合研究」のルーブリックがどのように作成され、その評価結果がどのように収集されてきたのかについて述べる。まず、2018年度当時の教養学部長であられた水谷修教授の指導のもと、教養学部教務委員会総合研究部会（以下、「総合研究部会」。当時の部会長は津上誠教授）が中心となり、「2019年度『総合研究』ルーブリック」（図1）が作成され、2019年3月18日の教養学部教授会でその原案が承認された。

⁴ 中教審（2012: 2-3）にもあるように、「学士課程教育の質的転換」は、さらにさかのほれば、中教審（2008）における「学士課程教育の質保証」達成のために行われるものであると思われる。

ルーブリックを用いた成績評価に関する報告—教養学部「総合研究」の事例—

評価観点	評価観点の説明	A	B	C	D
問題提起	この研究でどんな事柄を明らかにしようとしているかが明示されているか。また、その事柄に関する先行研究や一般的な理解のされ方を検討することによって、研究の学術的意義や社会的意義を十分に説明できているか。	明示されている。研究の学術的意義や社会的意義を十分に説明できている。	明示されている。研究の学術的意義や社会的意義をある程度まで説明できている。	明示されているが、研究の学術的意義や社会的意義がよく示されていない。	明示されていない。あるいは、研究の学術的意義や社会的意義が示されていない。
諸資料の収集、整理、提示	諸資料(調査や実験から得られる一次資料や文献資料)の収集、整理、提示の仕方が、関連する学問分野で妥当と認められるばかりでなく、優れた工夫や着眼点を含んでいる。	諸資料の収集、整理、提示の仕方が、関連する学問分野で妥当であり、かつ優れている。	諸資料の収集、整理、提示の仕方が、関連する学問分野で妥当と認められる。	諸資料の収集、整理、提示の仕方が、関連する学問分野で妥当とは言い難い。	諸資料の収集、整理、提示が不十分である。
引用表示	引用部分について、それとわかるような表示があり、かつ、開始部分や終了部分の表示、要約か原文かの表示、直接的引用か間接的引用かの表示、正確な引用元情報の表示などに、厳密さがあるか。	引用表示のルールが厳密に守られている。	引用部分について、それとわかるような表示があるものの、厳密さが不足しているため、引用表示の機能がある程度損なわれている。	引用部分について、それとわかるような表示があるものの、厳密さがまったくなく、有効な引用表示がなされた論文とは言い難い。	引用部分について、それとわかるような表示がほとんど、あるいはまったく、なされていない。
論理展開、および結論の内容	提起された問題に対し、諸資料を提示しながら、専門分野を異にする読者にとっても説得性ある論理展開を通じて、結論づけを行っているか。また得られた結論には十分な学術的意義や社会的意義が認められるか。	専門分野を異にする読者にとっても説得性ある論理展開を通じて、結論づけを行っている。得られた結論には十分な学術的意義や社会的意義が認められる。	専門分野を異にする読者にとっても説得性ある論理展開を通じて、結論づけを行っている。	結論提示に到る論理展開の説得性が、やや不足している。	結論提示に到る論理展開の説得性が、大変不足している。
形式面での適切さ(読者への配慮)	誤字脱字、文法上の誤り、章末での振り返りや章冒頭での内容予告の欠如、図表の見にくさ、といった形式面に関連する問題がなく、読者への配慮が行き届いた、読みやすい論文であるか。	形式面に関連する問題がなく、読者への配慮が行き届いた、読みやすい論文である。	形式面に関連する問題は少なく、読者にとって読みにくさは特に感じられない。	形式面に関連する問題が散見され、読者にとって読みにくい論文である。	形式面に関連する問題が非常に多く、読者にとっては意味を読みとるのが困難な論文である。
形成的評価	教員から指導を受けながら研究を進めているか。発表会等においては、自らの研究の進展状況や成果を十分に説明しているか。研究を進める際の主体性や、発表会等における他学生の研究に対する貢献が、よく認められるか。	教員から指導を受けながら研究を進めている。発表会等においては、自らの研究の進展状況や成果を十分に説明している。研究を進める際の主体性や、発表会等における他学生の研究に対する貢献も、よく認められる。	教員から指導を受けながら研究を進めている。発表会等においては、自らの研究の進展状況や成果を十分に説明している。	教員から指導を受けながら研究を進めていない、あるいは、発表会等において自らの研究の進展状況や成果を十分に説明していない。	教員から指導を受けながら研究を進めておらず、かつ、発表会等においても自らの研究の進展状況や成果を説明していない。

図1 2019年度教養学部『総合研究』ルーブリック

このルーブリックでは6つの評価の観点の設定され、それぞれAからDの4段階に評価される。教養学部の「総合研究」は、実験などによる実証的研究からプログラムやアプリケーションの開発、アンケート調査、社会的・地理的現地調査(フィールドワーク)、数理科学的研究、人文・社会・自然科学における理論研究や文献研究、思想・哲学・芸術論研究、言語および文化の横断的・個別的研究、演劇・映像作品の分析や作成・実演といった幅広い内容を含みうるため、ルーブリックの評価の観点は、できるだけそのような幅広さを包含しうるように設定され、かつ、この6つの評価の観点に各チーム独自の観点を追加することは妨げないこととされた。さらに、本学の場合、各科目の最終的な成績は0点~100点の点数で提出されるが、ルーブリックの評価をどのように最終成績の点数化に利用するか(たとえば6つの評価の観点を均等に点数化するか、どれかの割合を大きくするか、など)については、各チームの方針に委ねられている。

次年度の2019年度よりこのルーブリックの試用が始まり、評価結果の収集を開始した。2019年度は、各学生に対するルーブリック評価をプリントアウトした用紙に記したものの、またはExcelファイルに入力したデータなどを介して収集した。各学科における収集作業は総合研究部に代行していただいた。また、用紙に記したものについては、事務職員の助力を得てデータ化を行った。本格的運用の初年度となった2020年度も同様の方法で収集・データ化を行った。

本論での報告対象となる2021年度は、評価結果をGoogle Formを介して収集した。以下ではその概要について述べるとともに、収集されたデータについて簡易的な分析を提示する。

3. データの概要と分析

本節では、2021年度のルーブリック評価結果のデータ収集の概要を述べた後、学科および入試種別ごとの傾向の概要、学生の学修成果、評価結果と最終的な「総合研究」の成績との関係について、簡易的な分析を提示する。評価結果は、2022年2月に、上述のとおりGoogle Formで収集を行った。収集したのは、学生番号、各観点の評価結果（つまり、その学生に対する、各観点における4段階（A～D）の評価結果）、ルーブリックに対する教員のコメントである。前述の通り、2021年度、教養学部所属した教員総数は87名であり、うち、78名が「総合研究A」または「総合研究B」を担当した。その78名のうち、登録上、学生の所属があった68名のうち、62名による教員から、「総合研究B」に登録の学生451名中419名分の評価結果が提出された。よって、収集率は92.9%（=419/451）だった。「総合研究の遂行を放棄」「論文未提出」などの旨の記載があるものは全ての観点でD評価とみなした。A～Dの評価をA=3, B=2, C=1, D=0として数値化し、3点満点で学科（人間科学科・言語文化学科・情報科学科・地域構想学科）ごとの平均値を算出したものが表1である⁵。

⁵ FD研修会での報告の際、表1の人間科学科において、数値に若干の誤りがあったため、本論ではそれを修正してある。また、言語文化学科のある学生の評価に関して1つの観点のみ空欄の評価の提出があり、FD研修会報告時点では0として数値化したが、その後、評価結果を確認できたので、数値が若干更新されている。

表1 2021年度「総合研究」学科別ルーブリック評価結果

		観点						
		問題提起	諸資料の収集、 整理、提示	引用表示	論理展開、 および結論 の内容	形式面での 適切さ（読者 への配慮）	形成的評価	全体（6観点） 平均
学科	人間 [110]	2.16	2.07	2.49	1.87	2.08	2.23	2.15
	言語 [96]	2.17	1.82	2.01	1.66	2.07	2.11	1.98
	情報 [104]	2.29	2.09	2.15	1.97	2.09	2.25	2.15
	地域 [109]	2.52	2.39	2.32	2.34	2.33	2.51	2.40
全体 [419]		2.29	2.10	2.25	1.97	2.15	2.28	2.18

※ [] 内はデータ個数

表1からは、学生全体の値のうち、「論理展開、および結論の内容」の観点の値が1.97であり、他の値と比較して低いことがわかる。このことは、東北学院大学点検・評価委員会 編（2021: 50-51）で2019年度の評価結果に対して行われた分析でも指摘された傾向であると言える。また、学科ごとの評価結果としては、言語文化学科の「全体（6観点）平均」が1.98であり他学科と比較して低い数値になっている一方で、地域構想学科は2.40であり、逆に高い数値になっている。

入試種別ごとの評価結果からは、いくつかのことがわかり、FD研修会でそれらについて報告したが、ここではその概要についてのみ触れる。まず、「全体（6観点）平均」には、入試種別ごとの違いがみられたものの、データ個数が少ないものもあり、入試種別ごとの持続的な傾向かどうかを判断するには継続した調査が必要である。また、同じく「全体（6観点）平均」で比較すると、「アドミッションズ・オフィス（AO）による入試（A日程）」は2.27 [41]、「学業成績による推薦入試」は2.25 [94]、「TG推薦入試〔本学併設校からの推薦入試〕」は2.16 [46]、「一般入試（前期日程全学部型）」は2.22 [73]、「一般入試（前期日程学科分割型）」は2.15 [67]、「大学入試センター試験利用入試（前期日程）」は2.16 [33]であり、概ね同等の値であったことがわかる⁶。さらに、概ね同等の数値であるこれらの中でもAO入試と推薦入試によるグループの方が、一般入試とセンター利用入試によるグループよりも平均値が高かった。

ここまで、評価結果を学科ならびに入試種別ごとに分類し、そこから観察できる事柄について述べた。次に、「学修成果の可視化」という観点から、「総合研究」で身につけるべき能力を学生がどの程度獲得したのかを見てみたい。東北学院大学点検・評価委員会 編（2021: 50-51）で2019年度の評価結果に対して行われた分析によれば、当時、ルーブリックの評価を受け、そのデータが回収された教養学部生は「論理展開、および結論の内容」を除く各観点を、8割

⁶ ここでの入試種別名称は2018年度入試のものである。また、[] 内の数値はデータ個数である。

表2 2021年度「総合研究」ルーブリック各観点の評価の構成比率

			評価					
			A	B	C	D	合計	
観点	問題提起	個数	192	171	41	15	419	
		比率	45.8%	40.8%	9.8%	3.6%	100.0%	
	諸資料の収集、整理、提示	個数	138	202	62	17	419	
		比率	32.9%	48.2%	14.8%	4.1%	100.0%	
	引用表示	個数	188	156	48	18	410	
		比率	45.9%	38.0%	11.7%	4.4%	100.0%	
	論理展開、および結論の内容	個数	107	215	74	23	419	
		比率	25.5%	51.3%	17.7%	5.5%	100.0%	
	形式面での適切さ (読者への配慮)	個数	153	192	56	18	419	
		比率	36.5%	45.8%	13.4%	4.3%	100.0%	
	形成的評価	個数	209	141	47	22	419	
		比率	49.9%	33.7%	11.2%	5.3%	100.0%	
	全体 (6 観点)		個数	987	1077	328	113	2505
			比率	39.4%	43.0%	13.1%	4.5%	100.0%

前後がAまたはBの評価を得ている。同様の分析を今回の評価データで行った結果を表2に記す。なお、「引用表示」の観点についてのみ、合計が410となっているが、これは分野の慣例に従うと本ルーブリックでの評価ができない旨の評価が9件あったためである⁷。

表2からもわかるように、AとBの評価を合計した個数の割合は、7割台にとどまった「論理展開、および結論の内容」の観点を除けば、いずれも8割を超えている。このことから、多くの学生が「総合研究」で修得すべき能力を身につけたと評価されていると考えて良いものと思われる。

最後に、学生個人の評価結果と「総合研究」の最終評価（100点満点の成績）との関連について少し述べたい。前述の通り、このルーブリック評価をどのように最終成績に点数化するかは、各チームの方針による。しかしながら、たとえばルーブリックの評価が高ければ、最終成績も高くなり、逆に低ければ最終成績も低くなる、すなわち、両者は正の相関を持つと予測される。そこで、「全体（6 観点）平均」の値と「総合研究B」の成績として提出された100点満点の点数の関係を簡易的に調べた。そのために作成したものが図2である⁸。なお、ここで、「総合

⁷ ちなみに、本論では、これら9件の受講者は合計6観点ではなく合計5観点で平均値を算出している。また、表1の観点「引用表示」の平均値算出にあたっては、これら9件を除外している。

⁸ データ提出のあった419名分のうち、「総合研究B」の成績として、履修の「放棄」と登録した件数が7件あった。本学では、学務上、「放棄」と0点を区別しているが、この区別は今回の議論に関係ないため、図2においては、「放棄」を0点に換算している。

研究A」の成績を考慮せず「総合研究B」の数値を利用したのは、総合研究の成果物を提出するのが通常は後期であり、総合研究全体の評価およびルーブリックの評価がより直接的に反映されているのが「総合研究B」であると考えたためである。また、図2ではx軸が「全体（6観点）平均」（3点満点）、y軸が「「総合研究B」の点数」（100点満点）となっている。なお、y軸がマイナスの領域を含んでいるのは、後述の標準偏差を表示するためである。

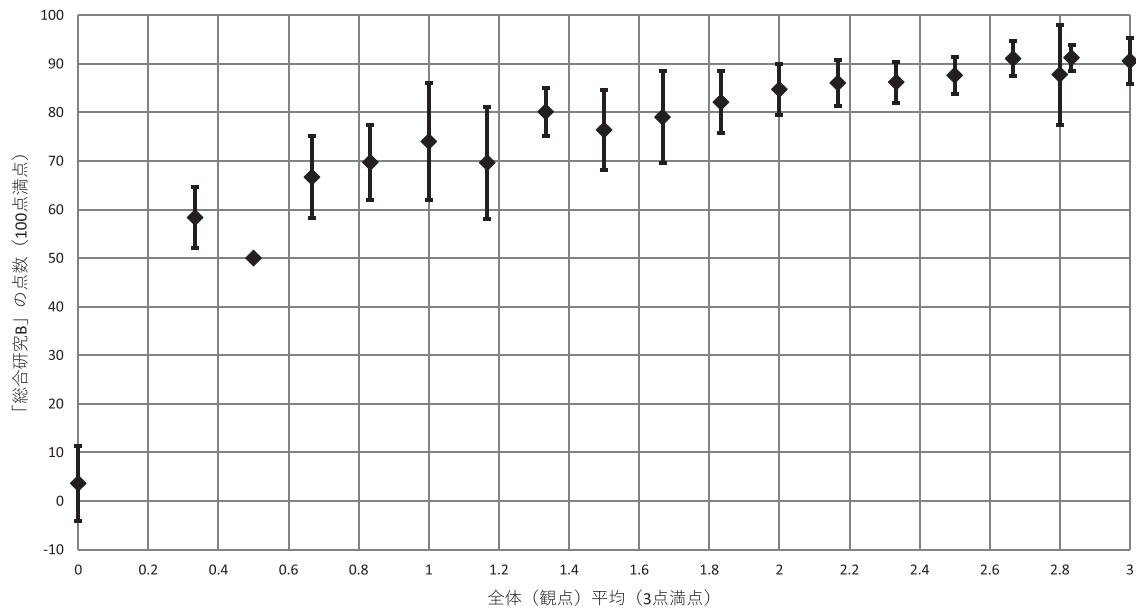


図2 2021年度「総合研究」ルーブリック評価結果と「総合研究B」の成績の関係

◆マークは、該当する「全体（6観点）平均」の受講者に対して与えられた「総合研究B」の成績の平均点の位置を示し、その上下に伸びる線は標準偏差の範囲を示している。（その範囲が示されていない◆マークが1つあるが、それは、標準偏差が0になるためである。）なお、ルーブリックの評価と「総合研究B」の成績との間に大きな乖離が見られるデータが2件（具体的には、前者が0であるにもかかわらず合格点が与えられたケース、ならびに前者が2に近いにもかかわらず不合格（0点）だったケース）含まれていたが、そのようなデータは例外的な値とみなして良いと思われるため、この平均値と標準偏差の算出からは除外してある。この図2を見れば、ルーブリックの評価が高かった受講者は「総合研究B」で高い点数を与えられ、逆の場合は低い点数を与えられる傾向にあることがわかる。さらに、標準偏差の幅を見れば、ルーブリックの平均値で概ね1.8を超える評価を受けた受講者は「総合研究B」での評価にばらつきが小さく、平均値に近い点数を付与される傾向があるが、それ未満の評価の場合は、「総合研究B」での評価のばらつきが大きく、平均値からより遠い点数を付与される傾向があることが読み取れる。

4. おわりに

以上、本論では、教養学部における、2021年度「総合研究」のルーブリック評価結果に基づき、学科および入試種別ごとの傾向の概要、学修の達成度、100点満点の成績との関連について述べた。このルーブリックに関しては、今後の改善のための課題として以下の点があげられるだろう。まず、第1に、評価A～DにSを追加した5段階評価への変更を検討することである。それにより、4点満点の点数化が可能になり、今回は着手しなかった、本学で導入しているグレードポイント（Grade point）とあるいはその科目平均であるグレードポイントアベレージ（Grade point average = GPA）との精密な比較ができるようになることが期待される。第2に、前期版のルーブリック開発に着手することである。現行のルーブリックは総合研究全体についての評価に関するものではあるが、前期の「総合研究A」での利用が難しい。なぜなら、6観点全てを「総合研究A」の評価に使用するのは困難であり、通常はごく一部のみを利用したものにならざるを得ないからである。

また、Google Formでのデータ収集のさいに寄せられた教員からの意見としては、独自のルーブリックとの整合性に関する意見や、このルーブリックで対応しきれていない部分についての意見が寄せられた。それらの意見も合わせて、今後、このルーブリックの加筆・修正を続けて行く必要があるだろう。

謝辞

本論の原稿に目を通し、多くの助言やコメントをくださった、水谷修先生、塚本信也先生、加藤健二先生、牧野悌也先生、津上誠先生、神林博史先生、高橋信二先生に厚くお礼申し上げます。特に、加藤健二先生には学務データを、牧野悌也先生には入試データをご提供いただく際、お世話になりました。また、総合研究部会の先生方にはルーブリック評価データ収集にご協力いただきました。この場をお借りし、お礼を申し上げます。言うまでもなく、本論のいかなる不備や誤りも、全て筆者一人の責任によるものです。

参考文献

スティーブンス、ダネル／アントニア・レビ（佐藤浩章（監訳）・井上敏憲・俣野秀典（訳））（2014）『大学教員のためのルーブリック評価入門』玉川大学出版部。

中央教育審議会（2008）「学士課程教育の構築に向けて（答申）」中央教育審議会。

(https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1217067.htm 2022年10月31日閲覧)

ループリックを用いた成績評価に関する報告—教養学部「総合研究」の事例—

中央教育審議会（2012）「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へ～（答申）」中央教育審議会.

(https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325047.htm 2022年10月31日閲覧)

東北学院大学教養学部（2017）「教養学部『教学上の方針』」東北学院大学教養学部.

(<https://www.tohoku-gakuin.ac.jp/faculty/liberalarts/about/policy.html> 2022年10月31日閲覧)

東北学院大学点検・評価委員会 編（2021）『2020年度東北学院大学点検・評価報告書』東北学院大学.

(https://www.tohoku-gakuin.ac.jp/about/evaluation/pdf/self_2020.pdf 2022年10月22日閲覧)