

大学生の協同学習を評価するルーブリック試作版の考案：看護学生を対象とした信頼性・妥当性の検討

Trial version of a rubric for university students participating in cooperative learning: An examination of reliability and validity among nursing students

會田 信子 Nobuko Aida

信州大学 学術研究院 医学保健学域 保健学系 Institute of Health Science, Academic Assembly School of Medicine and Health Sciences, Shinshu University

松井 瞳 Hitomi Matsui

信州大学 学術研究院 医学保健学域 保健学系 Institute of Health Science, Academic Assembly School of Medicine and Health Sciences, Shinshu University

加藤 善子 Yoshiko Kato

信州大学 学術研究院 人文社会科学域 総合人間科学系(高等教育研究センター) Research Center for Higher Education, Institute of Humanities, Academic Assembly School of Humanities and Social Sciences, Shinshu University

2017年7月24日投稿, 2018年4月17日受理

要旨

【目的】協同学習に参加している大学生のチームング・プロセスを評価するルーブリック TEAM-P2016v 試作版を考案し、信頼性・妥当性を検討した。【方法】対象学生は看護学生2年生69名で、評定者は3名だった。試作版ルーブリックによる観察法は、2016年8月にグループ探求法の協同学習場面において実施した。学生に対しては、無記名質問紙法を実施した。【結果】評定者間信頼性の級内相関係数0.792、一般化可能性係数(G係数)0.8197、内的整合性のカテゴリー主成分分析による信頼性 α 係数0.882、相関行列係数0.473-0.840で基準値を満たしていた。基準関連妥当性は、参考値として、学生の自己採点と他尺度との相関係数(0.241-0.413)を確認した。【結論】基準関連妥当性で課題は残されたものの、TEAM-P2016v 試作版は、チームング・プロセスのパフォーマンス評価指標として適用可能と考えられた。

Abstract

PURPOSE: We devised and examined the reliability and validity of a trial version of the TEAM-P2016v rubric designed to evaluate the teaming process of university students participating in cooperative learning. METHODS: The subjects were 69 second-year nursing students, who were evaluated by three raters. The observational method using the trial rubric was implemented in group exploration-based cooperative learning settings in August 2016. The students were asked to complete an anonymous questionnaire. RESULTS: The inter-rater reliability intraclass correlation coefficient was 0.792, the generalizability coefficient (G coefficient) was 0.8197, the alpha coefficient of reliability found by categorical principal component analysis of internal consistency was 0.882, and the correlation matrix coefficient ranged from 0.473 to 0.840, all of which met reference values. The criterion-related validity was used as a reference value to confirm the correlation coefficient between student self-evaluation scores and other scales (0.241-0.413). CONCLUSIONS: Despite residual issues with criterion-related validity, the TEAM-P2016v trial version was considered applicable as an evaluation index of teaming process performance.

キーワード

チームング、パフォーマンス評価、ルーブリック、協同学習、大学生 / 学士課程学生、信頼性・妥当性

Key Words

Teaming, performance evaluation/performance assessment, rubric (grading rubric), cooperative learning, undergraduate students, reliability & validity

1. 背景

今日の学士課程教育に対して、答えのない問題を自ら発見・解決しながら実行し、世界に通用する人材育成が求められている。具体的には、中央教育審議会から、学士の学位を保障する能力として「知識・理解」「汎用的能力」「態度・志向性」及び「総合的な学修経験と創造的思考力」が示され（中央教育審議会 2008）、それらの能力獲得を具現化するために、アクティブ・ラーニング(Active Learning; 能動的学修)（中央教育審議会 2012, 文部科学省 2012）の積極的導入が期待されている。アクティブ・ラーニングは、教員との双方向的な授業や、学生同士が刺激し合いながら思考・表現を引き出し、知性を鍛え、学生が能動的に学習し、主体的に成長していける課題解決型の学習法の総称である。

学習プロセスや獲得していく能力が従来の座学中心の一方的教授方法とは異なるため、教育評価方法も正誤・多肢選択問題や短答問題などのアチーブメント・テスト (achievement test) のみでは不十分で、学生の思考プロセスやコミュニケーション、技能、態度などのコンピテンス (competence; 能力) を可視化し、多面的な評価によって、更に学生のコンピテンスを涵養していけるパフォーマンス評価の意義が示されている（松下 2012, スティーブンス&レビ 2014, 田中 2010）。すなわち、パフォーマンス評価とは、「ある特定の文脈のもとで、様々な知識や技能などを用いて行われる人のふるまいや作品を、直接的に評価する方法」（松下 2007）で、学生に期待される成果基準にもとづいて、直接的・体系的に観察して評価する。そして、パフォーマンス評価の信頼性を保障するための指標として、教育上の到達点の内容が評定尺度として設定されたルーブリック (rubric) が必要となる。また、そのルーブリックを学生にも公開し共有することで、学生の学習活動上のペースメーカー的役割や自己評価の指針となり、学生自身にも有用といわれている（松下 2007, 田中 2010）。

既存のルーブリックには、学位授与方針の能力 (American Association of Colleges & Universities n.d., The Council for Aid to Education n.d.) や、小論文・レポートを評価するもの（松下 他 2013, 宇佐見 2011）、模擬場面を設定して行う学生

の多職種協働力 (interprofessional collaborator competencies) や臨床推論・判断力を評価するもの (Curran et al 2011, Lasater 2007, Shin et al 2015, Yeung et al 2016) があるが、今回、我々は、協同学習 (cooperative learning) におけるパフォーマンス評価のためのルーブリックを開発する。

協同学習とは、(1) 互恵的相互依存関係の確立 (目標達成にはすべてのメンバーによる相互協力が不可欠である)、(2) 二重の個人責任の明確化 (学習者個人とグループ全体の学習目標を達成することを構成員が承知して参加する)、(3) 促進的相互交流の保障と顕在化 (学習目標を達成するために構成員による相互協力のもと実施する)、(4) 協同の体験的理解の促進 (協同の価値・効用の理解・内化の促進のために教師が意図的に働きかける) の4つの条件を満たす、もしくは、満たそうと意図される2名以上のグループ学習であると共通認識されている (バークレイ 他 2009, ジョンソン 他 2010, 関田・安永 2005)。このような協同学習の4つの条件からも、協同学習の教育評価においては、課題の成果物 (作成された資料・発表内容、個人レポートなど) のコンテンツ (content; 内容) に焦点を当てることは勿論であるが、協同学習の目的ともいえる4条件を反映する学習プロセスについても評価視点として加え、ルーブリックで評定していく必要性が考えられた。

その背景として、協同学習中の学生のいわゆる「社会的手抜き」(ウェスト 2014) などの問題があげられる。社会的手抜きは、グループ内で一人でも努力の手抜きをする者がいると、他のメンバーは、自分達が利用され、ただ乗りされたと感じ、満足感や達成感が低下してしまうことが指摘されている (ウェスト 2014)。また、グループ学習中の意見交換の不活発さや作業停滞などが観察されながらも、高いレベルの成果物が提出される場合は、特定メンバーの個人作業による場合が考えられ、教員の役割として、アクティブ・ラーニングの本来の目的との乖離を解決していくことが必要である。看護学生を対象に『協同作業認識尺度』を用いた調査でも (會田 他 2017)、協同作業は学習方法として有効であるとの《協同効用》の認識が低く、個人作業を好む《個人志向》が高く、なおかつ協同作業に対する《互恵懸念》(協同作業はお互いが等しく恩恵を受けることが常に保証されてい

るわけではなく恩恵は少ないとの認識)が高い看護学生の割合は約3割みられた。そして、これら3割の学生は、そうでない学生と比較して、他者の能力を軽視する傾向のある仮想的有能感が高く、ソーシャルサポート得点が低いなどの結果が確認されたことから、問題解決に必要な協同作業を遂行していく力を客観的に評価し、学生の成長につなげていくことが重要と考えられた。

しかしチームワークを評価する既存の日本語版尺度には、協同作業に対する認識(長濱 他 2009)や、学生が自分のチームワーク力を自己評価するもの(相川 他 2012, 三沢 他 2009, 山口 2007)、所属チームのチーム力を主観的に評価するもの(飯岡 他 2016, 池田・古川 2009)、もしくはチーム内の共感性や協調性を自己評価するもの(阿部 他 2012, 小味 他 2011)など、ほとんどが個人の認識や能力を自己採点するもので、第三者が学生のパフォーマンスを評価するルーブリックは開発されていない。

国外の文献において、模擬患者に対する多職種チームのチームワーク行動(teamwork behaviors)を測定する尺度が精力的に開発されており、ビデオ撮影をして数値化するCATS(The Communication and Teamwork Skills; Frankel et al 2007)やMHPTS(The Mayo High Performance Teamwork Scale; Malec et al 2007)、CTS(The Clinical Teamwork Scale; Guise et al 2008)、TPOT(The Team Performance Observation Tool; Agency for Healthcare Research and Quality 2014)などがある。また、チーム・パフォーマンス(team performance)をキーワードとした尺度には、医学・看護・理学などの学部生を対象に、多職種チームによる模擬課題を設定して、メンバーとリーダーの役割遂行のチーム機能を評価するチェックリスト(KidSIM Team Performance Scale Checklist; Sigalet et al 2013)が開発されている。これらは、チームを1つの単位として評価していたり、学生個々を評価していてもルーブリックの形式でなかったりなど、本研究が目指す協同学習中の学生個人に焦点をあてたルーブリックは海外でも開発されていない。

そこで、今回は、Edmondsonのチーミング(teaming)(Edmondson 2012, エドモンドソン 2014)と津村のプロセス・エデュケーション(津

村 2012)の概念モデルを基盤として、協同学習に参加している学士課程の学生用(以下、大学生とする)のチーミング・プロセスを教員が評価するためのルーブリック TEAM-P2016v(Grading Rubric of Teaming-Process Performance 2016 version)の試作版を考案し、B大学の看護学生を対象として信頼性・妥当性を検討した。

2. 目的

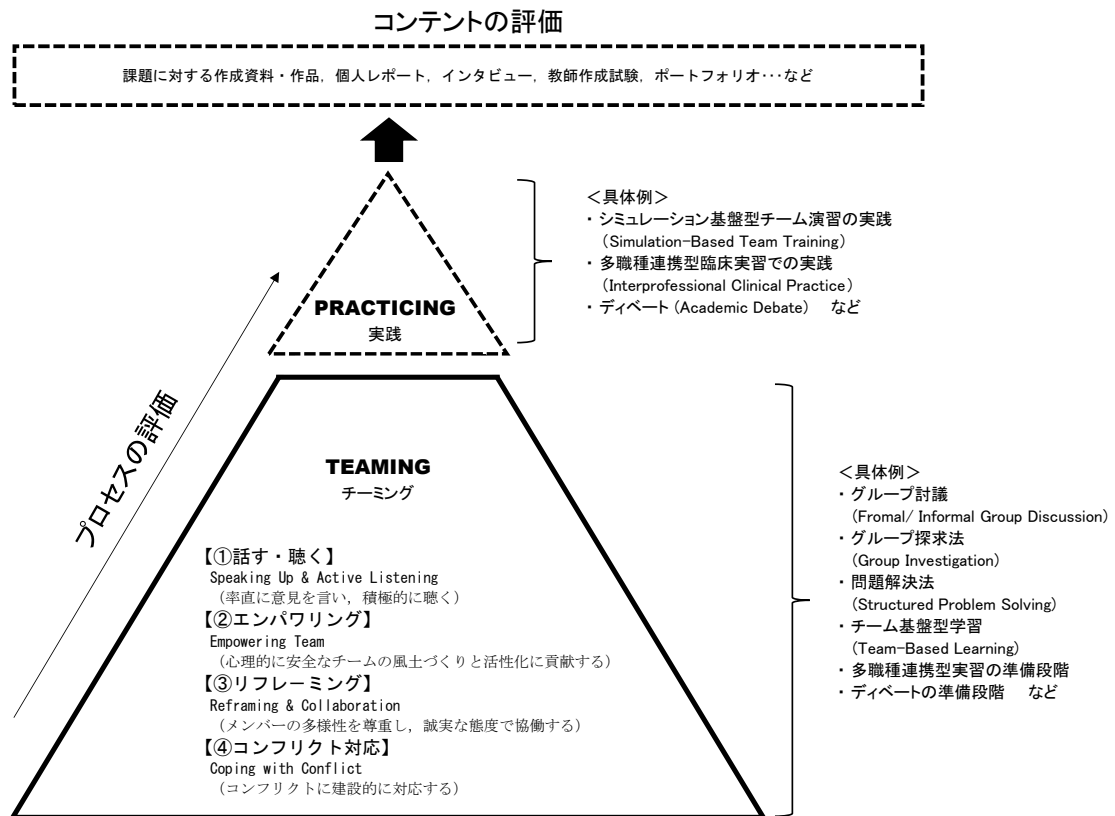
本研究の目的は、ルーブリック TEAM-P2016v 試作版を考案し、信頼性・妥当性を検討することで、具体的目標は下記の3つである。

- 1) 協同学習に参加している大学生用ルーブリック TEAM-P2016v の試作版を考案する。
- 2) 協同学習に参加している看護学生を対象として、ルーブリック TEAM-P2016v 試作版の評定者間信頼性、項目の集約可能性、内的整合性および基準関連妥当性を確認し、大学生のチーミング・プロセスを評価するためのパフォーマンス評価指標としての適用の可能性を検討する。
- 3) 評定者と学生のルーブリック TEAM-P2016v 試作版の得点を比較し、教育評価のあり方への示唆を得る。

3. 本研究の概念枠組み

3.1 概念枠組みを考案するにあたり基盤とした理論
ルーブリック TEAM-P2016v 試作版を考案するに当たっては、Edmondsonの「チーミング(teaming)」の考え方(Edmondson 2012, エドモンドソン 2014)と、津村(2012)の「プロセス・エデュケーション」の教育モデルを基盤として、図1に示す概念枠組みを独自に作成した。

チーミング・プロセスの「チーミング」はEdmondsonの造語で、協働するという「活動」を表す『動詞』として、チームワークの根底、組織学習の原動力をなすものとしている。Edmondsonの二十数年にわたる病院や製品開発チーム、工場の組み立て部門などのスタッフや管理職を対象にした実施調査からまとめられた概念で、協働する可能性がある瞬間に、時間や場所を選ばずに学習する力、実行できる力を兼ね備えたチームを機能させるプロセスや考え方などをまとめたものである。Edmondsonによれば、従来のチーム理論



ルーブリックTEAM-P2016v (Grading Rubric of Teaming-Process Performance 2016 version)

本図の中の<具体例>は, 医療系大学で実施されている協同学習 (cooperative learning) の教育・評価方法を研究者の経験や論文などから抽出して示した。実線の台形内のチーミング/TEAMING部分が, 本研究のルーブリックTEAM-P2016vでカバーする内容を示す。

図1. 本研究における概念枠組みとルーブリックTEAM-P2016v試作版がカバーする内容(実線部分)

が, 単に「チームを作るための学習」に留まり(名詞のTeam)、社会的に有意義な成果に注目していない弱点をあげ, 新たなアイデアを生み, 答えを探し, 問題を解決するために人々を団結させる働き方や協働のあり方が必要と強調し, その基盤となるプロセスや考え方を「チーミング(動詞としてのTeaming)」の概念として理論化したものである(Edmondson 2012, エドモンドソン 2014)。

本研究で扱う「協同学習」の方法も, 緒言で述べたように, チームを作ること自体が最終的な教育目標ではなく, 学習者個人が明確な責任を自覚しながら, 全メンバーによる相互交流や協力を図り, グループに課せられた学習目標を達成し, 協同の価値・効用の理解を通して成長を促していくことを目指している(バークレイ 他 2009, ジョンソン 他 2010, 関田・安永 2005)。確かに, チーミング理論は, 企業や病院などの組織における実地調査から生成された理論であるが, 協同学習の条件が, 学生の「チームの特性」に焦点を当ててい

るのではなく, 課題を達成するための学生同士の「働き方」の学習プロセスに重きを置いていることから, チーミング理論と共通する視点があると考え, Edmondsonのチーミング(teaming)の考え方(Edmondson 2012, エドモンドソン 2014)を参考とした。

津村のプロセス・エデュケーションとは, 「教育者と学習者が, 学習者と学習者との関係を築くこと」で, 外部から観察できるゾーンとして, コンテント・ゾーン(やった仕事や課題)と関係のプロセス・ゾーン(集団の形成・維持)の2つを示している。今回, 津村のプロセス・エデュケーションのモデルを採用したのは, ルーブリックTEAM-P2016v試作版が評価する内容は「コンテント・ゾーン」ではなく, 「プロセス・ゾーン」であることを明確にすることが重要と考えられたためである。またルーブリックTEAM-P2016v試作版がプロセス・ゾーンに焦点を当てて, コンテント・ゾーンを評価から除外したのは, コンテント

の評価内容や方法（学生が作成した資料や個人レポートなど）は、教育科目の「単元」や「課題内容」に左右され多種多様になることが考えられたためである。従って、本研究では、どのような単元や課題内容にも共通するプロセス・ゾーンのチーミング・プロセスに焦点を当ててルーブリックを考案することとした(図1の概念枠組みの実線部分)。

3.2 下位概念の検討

概念枠組みの下位概念に相当する【項目】を抽出するにあたっては、Edmondsonのチーミング理論から質的記述的に行った。最初に、Edmondsonのチーミング理論のモデルをもとに、関係する用語をすべてエクセルに抽出した後、概念間の関係性や意味、抽象度などをもとに用語の整理を行った。その結果、チーミングモデルで多用されている用語として、「優れたコミュニケーション」「率直に意見を発言する」「他者の意見を取り込む」「質問する」「意見を求める」「助けを求める」「提案する」「他者を受け入れる」「省察」「協働」「尊敬し合う」「誠実」「信頼」「緊張と安心」「対人リスク不安」「目標の共有」「チームの雰囲気」「帰属意識」「他者理解」「協調的取り組み」「差異への認識」「意見の不一致」「対立への建設的対応」「自己調整」「論理的思考」「合意形成」などの45個が抽出された。

次に、協同学習の目的・考え方、必要とされる技能、および実施上の留意点などの内容を協同学習に関連した図書(バークレイ 他 2009, ジョンソン 他 2010, 関田・安永 2005)を参考にしながら抽出した。その結果、「積極的参加」「話し合い(ディスカッション)」「自己開示」「簡潔な発言」「積極的傾聴」「アサーティブ」「雰囲気づくり」「対人技能」「礼儀作法」「挨拶」「自己紹介」「民主的態度」「他者への思いやり」「他者への敬意」「対等な関係性」「相互協力」「相互交流」「主体性」「個人の責任」「目標の共有」「意見のズレ」「対立への建設的対応」「暗黙の解釈」「認知フレーム」「学習のためのリフレーミング」「自他ともの成長」などの35個が抽出された。

上記のチーミング理論と協同学習から抽出された用語から、両者に共通する内容を絞り込んだ後、これらの類似性と相違性から質的記述的に分類作業を行った。分類にあたっては、Edmondsonの

書籍の章立てで使用されている用語の抽象度の上下関係や、本文での用語の意味や概念の関連性を意識しながら4つにカテゴリ化して、それらを代表する内容を解釈して命名した。

その結果、①率直に意見を言い積極的に聴く(Speaking Up & Active Listening) (以下【①話す・聴く】とする)、②心理的に安全なチームの風土づくりと活性化に貢献する(Empowering Team) (以下【②エンパワリング】とする)、③メンバーの多様性を尊重し誠実な態度で協働する(Reframing & Collaboration) (以下【③リフレーミング】とする)、④コンフリクトに建設的に対応する(Coping with Conflict) (以下【④コンフリクト対応】とする)とした。なお項目の【 】内の名称は、本文内で略語的に表記するために研究者が便宜的に作成した。

4. ルーブリックTEAM-P2016v試作版の考案

4.1 観察するパフォーマンスの考案

考案したルーブリックTEAM-P2016v試作版は、表1に示した。

概念枠組みの4つの【項目】の観察内容となるパフォーマンスを検討するにあたっては、学士課程の協同学習で共通的に期待される内容を提示する必要があることから、以下の4つを参考資料として活用した。すなわち、(1)研究者が過去に学部教育で行ってきた協同学習を取り入れた単元のシラバスに記載した評価の視点、(2)グループワークをテーマとした先行調査に記述されている期待される言動や評価の視点(藤野 2005, 芳我・舟島 2007, 市川 他 1994, 小山 1995, 国澤 他 1996, 松田・牧野 2012, 塚崎 1996)、(3)高等教育機関の協同学習関連の書籍等に示された、学生に期待される言動の内容(バークレイ 他 2009, ジョンソン 他 2010, 関田・安永 2005)、および、(4)チームワーク力などを測定する尺度の質問文(相川 他 2012, 飯岡 他 2016, 池田・古川 2009, 三沢 他 2009, 長濱 他 2009, 山口 2007)である。

具体的な作成手順は、最初に、前段落の参考資料(1)(2)(3)(4)から抽出したすべての内容をエクセル・ファイルにすべて抜き出した後(1つのセルに1つのパフォーマンス内容)、Edmondsonのチーミング理論で重要としている前提や考え方

表1. 考案したルーブリックTEAM-P2016v試作版

	評定段階			
	キャップストーン 3	マイルストーン 2	ベンチマーク 1	ベンチマーク未満 0
【①話す・聴く】 Speaking Up & Active Listening (率直に意見を言い、積極的に聴く)	以下のすべてを行うことで、チームメンバーと基本的なディスカッションができる。	以下のうち3つ行うことで、チームメンバーと基本的なディスカッションができる。	以下のうち1~2つ行うことで、チームメンバーと基本的なディスカッションができる。	ベンチマーク・レベルのパフォーマンスを満たさない。
	1) 自分の意見・情報を提示・提案するとともに、メンバーに質問・意見を求める。 2) メンバーの反応に意識を向け、発言や質問を聴く。 3) ディスカッションにふさわしい話し方（声の大きさ、強さ、速さなど）である。 4) 会話での言葉・発言内容と非言語的行動が一致している。			
【②エンパワリング】 Empowering Team (心理的に安全なチームの風土づくりと活性化に貢献する)	以下のすべてを行うことで、開かれた支持的・建設的な雰囲気を支え、ワークの活性化に貢献する。	以下のうち2つ行うことで、開かれた支持的・建設的な雰囲気を支え、ワークの活性化に貢献する。	以下のうち1つ行うことで、開かれた支持的・建設的な雰囲気を支え、ワークの活性化に貢献する。	ベンチマーク・レベルのパフォーマンスを満たさない。
	1) メンバーの意見・見識を尊重し、誠実な態度である。 2) 支持的・建設的な言語的・非言語的メッセージでチームの雰囲気を支える。 3) チーム機能の不全状態に対して、真摯な姿勢で客観的・中立的に働きかける。			
【③リフレーミング】 Reframing & Collaboration (メンバーの多様性を尊重し誠実な態度で協働する)	以下のすべてを行うことで、お互いの立場・意見を尊重しながら、幅広い多角的視点で一緒にワークをすることができる。	以下のうち2つ行うことで、お互いの立場・意見を尊重しながら、幅広い多角的視点で一緒にワークをすることができる。	以下のうち1つ行うことで、お互いの立場・意見を尊重しながら、幅広い多角的視点で一緒にワークをすることができる。	ベンチマーク・レベルのパフォーマンスを満たさない。
	1) 自分と異なる立場（属性・職種・価値観など）のメンバーに対して、否定的・攻撃的な言語的・非言語的メッセージを発することがない。 2) メンバーに対して、開かれた態度で、作業調整・支援、フィードバックを行いながらワークを進める。 3) 学習課題を自分とメンバーが同じレベルで遂行・達成できるように率先して協力する。			
【④コンフリクト対応】 Coping with Conflict (コンフリクトに建設的に対応する)	以下のすべてを行うことで、破壊的な対立を回避し、チームビジョンに向かって建設的な取り組みに貢献できる。	以下のうち2つ行うことで、破壊的な対立を回避し、チームビジョンに向かって建設的な取り組みに貢献できる。	以下のうち1つ行うことで、破壊的な対立を回避し、チームビジョンに向かって建設的な取り組みに貢献できる。	ベンチマーク・レベルのパフォーマンスを満たさない。
	1) 自分と異なる意見・価値観のメンバーに対して、感情的にならずに対応する。 2) 自身への批判的発言に対して、前向きに根気強く対処する。 3) 自身やメンバーのミスを受け入れ、失敗と有意義に向き合う言動を発する。			

ルーブリックTEAM-P2016v(Grading Rubric of Teaming-Process Performance 2016 version)

と照らし合わせながら、観察すべきパフォーマンス項目として絞り込んだり、あまり関係ない内容を削除したりしながら、項目を取捨選択していった。次に、絞り込まれたそれらの内容を、概念枠組みのどの【項目】に関係するかを内容の類似性・相違性の視点から分類した。分類されたパフォーマンス項目を、観察項目の文章として洗練化していくにあたっては、チーミング理論で重要視しているボキャブラリーを優先的に盛り込みながら40字以内で文章化した。

ルーブリックの素案は、研究責任者である著者が一人で作成した。内容妥当性を高めるために、

共同研究者に、普段、自身が担当する単元の協同学習場面をイメージしながら、内容に対する違和感や観察しづらい内容などがいないかを確認してもらい、指摘事項について修正を行った。この加筆修正作業は計2回実施した。

なお、本ルーブリックの評価内容から、既存のチームワーク尺度で見られる、より高いチームワーク力は除外した（例として、ブレイクスルー的発想を生み出すためのリーダーシップ力やパワーバランスの調整、他の組織との交渉力など）。その理由として、これらの発展的なチームワーク力は、専門職や企業スタッフを対象とした尺度

の測定項目に組み込まれていることから(飯岡 他 2016, 池田・古川 2009, 三沢 他 2009, 山口 2007)、学士教育課程の教育目標として一般的にひろく設定されていないと考えられたためである。

4.2 ルーブリックの評定段階

評定段階については、メンバー構成や課題内容などによって学習プロセス自体がダイナミックに変動する可能性を前提としていることから、ルーブリックTEAM-P2016v試作版では、教育的支援を必要とする学生の吸い上げに有効(松下 他 2013, 宇佐見 2011)と考えられる4段階評定(3 キャップストーン、2 マイルストーン、1 ベンチマーク、0 ベンチマーク未満・不可)を採用することとした。

たしかに5～7段階評定のルーブリックも散見される(American Association of Colleges & Universities n.d., The Council for Aid to Education n.d.)。しかし、5～7段階と細分化された評価指標は、学位授与に必要とされる学年ごとのスマール・ステップ的な成長を、学生の自己評価とともに長期間に渡って確認していくには優れているが、限られた時間内で評価するには不向きと考えられた。

また、本ルーブリックでは、3～4個の観察項目を設けて、そのうちの幾つできるかによって評定段階を区別した。すでに公開されている小論文などのルーブリックでは(沖 2013, 松下 2007, 松下 2012, 松下 他 2013, ステューブンス&レビ 2014, 田中 2010, 宇佐見 2011)、評価項目の内容を「模範的に～できる」「完璧に～作成されている」「最低限～できている」など、達成レベルを形容詞で示したり、「AとBはできている」「Aのみできている」「5個中1個ないし2個についてできている」など、到達している項目の可否や数によって評価したりしているものが多く散見される。

達成レベルを形容詞で示して評定する前者のルーブリックは、どのようなパフォーマンスがみられれば「完璧」なレベルといえるのかなど、限られた時間の中で評定者の迷いが生じやすいデメリットがある。また、どのようなパフォーマンスが「模範的」といえるのかを具体的なパフォーマンスとして記述しても、その記述内容に合致しない学生のパフォーマンスが観察されるため(評価の

対象とならない)、評価のフロア現象が生じる危険性が確認された。さらに、具体的なパフォーマンスを記述すると、ルーブリックに提示される説明文が多くなるため、評価に多くの時間を費やす課題も明らかになった。以上のことを加味した結果、本ルーブリックでは、4つの【項目】ごとに、それを構成する観察項目を複数提示し、そのうち幾つできているかで評定段階を評価する方法を採用することにした。

なお、ルーブリックTEAM-P2016v試作版の4段階評定の<レベル3>は、期待されたパフォーマンスが秀でて観察されるキャップストーンレベルを、<レベル3>には至らないが、期待されたパフォーマンス項目の内容が標準的に観察されるマイルストーンレベルを<レベル2>、そして<レベル2>には及ばないで不十分な項目があるベンチマークレベルを<レベル1>とした。最後の<レベル0>は、期待されるパフォーマンスが観察されないため、なんらかの教育的介入・支援を必要とする、いわばベンチマーク以下の合格水準に至らない「不可レベル(unable to evaluate)」とした。

5. 方法

5.1 対象学生

対象学生は、関東信越地方に所在するB大学の看護学生で、『老年看護学実習I』(前期・2単位)を履修している第2学年の70名であった。なお実習は、実習時期によって前半と後半の2グループに分かれており(各グループが2週間ずつ実施)、1グループの人数は35名であった。

後述するように(5.3 調査時期と手続き参照)、観察法は全対象の70名に実施したが、最終的にデータ分析の対象としたのは、調査協力に直筆で同意の得られた69名(うち男性10名、14.5%)であった。同意率・回収率は98.6%、有効回答率100%であった。

5.2 評定者

評定者は、看護系大学で教育経験を有する老年看護学実習を担当している2名(研究者および共同研究者)と、当該科目の教育担当経験のある1名の計3名(全員女性)で、事前に本研究の趣旨とルーブリック評定上の留意点などを共有して実施

した。

評定者数を3名にしたのは、評価の実行可能性を低下させない最少人数で(松下他2013)、なおかつ、カリキュラム上の教育課題や学生の学習プロセスに悪影響を及ぼさないことを考慮したためである。また評定者の教育経験年数などが類似していると、評定にバラツキが生じにくいデメリットがあるため、教育経験年数があえて異なる評定者3名を選定した。なお、評定者の教育経験年数は、それぞれ20年以上、5年以上10年未満、1年未満であった。

5.3 調査時期と手続き

調査時期は2016年8月で、前半・後半グループのそれぞれの実習期間中の実習3日目に、B大学の演習室で実施した。観察法は、教科目のグループ探求法の協同学習中に(詳細は後述の調査内容5.5.2(2)参照)、質問紙法は、協同学習終了後に実施した。なお、協同学習の課題内容は「実習施設の特徴・資源を踏まえたアクティビティ・ケアの計画立案」であった。

調査手続きは、最初に、その日の学習課題を説明した後、研究説明書を用いて研究の趣旨や方法などを説明した。その後、研究協力への同意書を直筆で記載してもらい(氏名、着席番号、同意の有無)、施錠付きのボックスに投函してもらった。学生には、協同学習中に全学生を観察するが(前半35名、後半35名)、同意しなかった学生については、着席番号で付された観察用紙と氏名を対応表で確認して、収集した観察用紙や質問紙をすべて破棄すること、質問紙にも回答しなくていいことなどを説明した。同意書を投函してもらった後、実習の課題に対する協同学習を実施してもらった。質問紙(着席番号を付記)は、協同学習終了後の実習時間中に記載してもらい、施錠付きのボックスに投函してもらった(記載にかかる平均所要時間5~10分)。

なお同意書の記載・投函時に、評定者3名は学生の視界に入らない場所に移動した。また質問紙については、学業に支障を及ぼさないよう、午前中の協同学習が終了した昼食休憩時間中に記載し、施錠付きボックスに投函してもらうよう協力依頼した。

5.4 倫理的配慮

調査は、信州大学医学部医倫理委員会で承認を得た後(承認番号3389)、学生の自由意思に基づいて十分な倫理的配慮のもと行った。収集したデータには個人情報とは無関係の「識別番号」を付けて、統計ソフトに入力し個人が特定できないようにした。同意の撤回に備えて個人を識別することができるように、同意書に記載してもらった「氏名」「着席番号」の情報を含んだ対応表をエクセルで作成した。対応表は、パスワード付きのディスクに保存した後、所属施設の倫理規定に定められた情報管理責任者によって、回収した観察用紙と質問紙と一緒に施錠付き保管庫で管理した。学生に対しては、統計解析時は、個人情報とは関係のない識別番号を用いた統計ソフトを利用すること、実習の成績評価には反映されないことなどを説明した。

質問紙の属性における「年齢」は、留年生など、個人が特定される可能性が考えられたため、質問項目から除外した。また、成果公開時は「関東信越地方に所在するB大学の第2学年の看護学生」と表記する旨を説明して同意を得た。

なお尺度の使用にあたっては、著者の責任において、著作権所有者を確認し、倫理委員会申請書や論文に出典を明記する対応をとった。

5.5 調査内容

5.5.1 観察法

観察法は、考案したルーブリックTEAM-P2016v試作版を用いて、グループ探求法による協同学習中に実施した。一人の学生に対して評定者3名が、別々にラウンドしながら観察して評定した。評定者が1回の協同学習で評価した学生数は35名であった。これは、すでにカリキュラムに組み込まれている実習期間を利用したため、前半・後半グループの2回にわたって35名ずつ実施した。(5.1対象学生と5.3調査時期と手続き参照)。なお、1回のグループ探求法の観察時間は100分であった。

観察法では、ホーソン効果(Hawthorne effect)が強まったり、協同作業の邪魔になったりしないよう、ルーブリックの観察項目の詳細内容は研究の説明時には提示しないで、「グループ学習の様子を観察させてもらう」との説明のもと実施した。

また、学生の着席位置から、最低でも1メートルは距離をおいて、学生の視界に可能な限り入らない場所で行った。また評定者間の話し合いによる評価は禁じ、各評定者が観察して評価した数値をそのまま記載してもらった。従って、個々の評定者は教室内を自由にラウンドしながら、所定の時間内で学生のパフォーマンスを評価した。

5.5.2 無記名自記式質問紙法

(1) 基本属性

対象の基本属性は、性別と学年の2項目であった。

(2) 高等学校での協同学習の経験

高等学校(以下、高校とする)での協同学習経験は、基準関連妥当性検証のための変数として用いた。本変数を用いた理由は、協同作業認識尺度の《個人志向》の下位概念の有意な影響要因であったことや(會田 他 2017)、高校での「探求学習活動」経験が「サービス満足」や「授業満足」などの『大学満足』への直接効果(パス解析)として確認されたこと(木村 2010)、さらに大学生のチームワーク能力を向上させるトレーニングを実施した群は、非実施群と比較して社会的スキルやモニタリング能力、遂行指導力が有意に高値であったとの調査結果(太幡 2017)から、協同で作業を行う経験やトレーニングが、ルーブリック得点に影響すると考えたためである。

協同学習経験は、協同学習の様々なスタイル手法を紹介した書籍を参考に(バークレイ 他 2009, ジョンソン 他 2010, 関田・安永 2005)、一般的に広く実施されていると考えられる教育手法、すなわち、(i) 授業中に近くの学生2~3名と組んで意見交換などをするインフォーマル・グループ・ディスカッション(以下、インフォーマルGDとする)、(ii) 学生2~3名で教科知識の理解度や技術の達成度を確認し合ったり、お互いに教え合ったりするシンク・ペア・シェア(think-pair-share; 以下、シンクペアSとする)、(iii) 学生4~6名がシンクペアSと同じ方法で行う学生同士の教え合い技法(以下、教え合い技法とする)、(iv) 学生が2名以上で課題に取り組んだり、資料などを作成したりするグループ探求法、(v) 学生が2名以上でプレゼンテーションを行い、クラス全体でディスカッションするストラクチャード・プロブレム・

ソルビング(structured problem solving; 以下、問題解決法とする)、(vi) ディベートの6つの技法を抽出して、その技法に準じた平易な設問文を提示した。選択肢は、4件法で回答を得た(3頻回に経験した、2時々経験した、1数回位は経験した、0一度も経験なし)。

(3) 個人のチームワーク能力

ルーブリックTEAM-P2016v試作版と他の心理尺度との整合性(基準関連妥当性; criterion-related validity)を確認するため、学生が自分のチームワーク力を振り返る『チームワーク能力尺度』(相川 他 2012)を使用した。多くの尺度が専門職や企業スタッフを対象として作成されているのに対し、本尺度は大学生用として開発されており、信頼性・妥当性も確認されていることから、隣接概念を測定する尺度として採用した。

質問内容は、個人がチームに属した時に発揮する能力を、普段の自分を振り返って回答するもので、<コミュニケーション能力>17項目、<チーム志向能力>13項目、<バックアップ能力>15項目、<モニタリング能力>12項目、<リーダーシップ能力>15項目の5つの下位尺度からなる。特定チームに所属していなくても回答できるよう、教示文に「あなたがあるチームに属していると仮定してお答え下さい」と示されていることから汎用性が高い。本尺度の選択肢は6件法からなり(1まったくあてはまらない、2ほとんど当てはまらない、3あまり当てはまらない、4少し当てはまる、5だいぶ当てはまる、6非常にあてはまる)、得点が高いほど(逆転項目はリバーズ処理)、その個人が、チームに属した時に発揮するチームワーク能力が高いと認識していることになる。

(4) ルーブリックTEAM-P2016v試作版を用いた自己評価

ルーブリックTEAM-P2016v試作版を用いて、協同学習中の自身のチーミング・プロセスを自己採点してもらった。これは、妥当性を検討するため、学生が自己採点したルーブリックTEAM-P2016v試作版と『チームワーク能力尺度』の得点が相関関係を有するかを確認するためである(後述分析方法5.6(4)参照)。なお、観察法と同様に、グループ探求法でのホーソン効果を可能な限り避けるために、質問紙のルーブリックと『チームワー

ク能力尺度』の内容は、研究の説明時には提示しないで実施した。

5.6 分析方法

全変数について記述統計を実施した後、コルモゴロフ・スミルノフ検定とシャピロ・ウィルクス検定で正規性を確認した。その結果、3名の評価者による TEAM-P2016v 試作版の合計点と平均値、および学生のチームワーク能力尺度の合計点で正規性が確認できなかったため下記の方法で実施した。

(1) 高校での協同学習経験の点数化

高校での協同学習経験は、各形態の経験頻度を点数化して(3頻回に経験した、2時々経験した、1数回位は経験した、0一度も経験なし)、それらを合算した合計点(以下、協同学習の経験頻度得点とする)をヒストグラムで視覚的に確認した後、中央値と平均値、範囲などを求めた。

(2) 評定者間信頼性

ルーブリック TEAM-P2016v 試作版の評定者間信頼性(intra-rater reliability)は、最初に3名の評定者の一致率の割合を確認した(全員不一致、2名が一致、3名が一致)。その後、3名の評価者が69名の学生に配点したルーブリック得点の評定者間比較を行うために、対応サンプルによるFriedmanの順位付けによる比較と多重比較(順位付けによるFriedmanの二元配置分散分析)を行った。その後、平均測定値の級内相関係数(intraclass correlation coefficients; 以下ICCとする)(二元配置混合ICC(3, 1))を確認した(基準値は0.70以上)。

さらに、分散成分分析を用いて信頼性を推定する一般化可能性理論(generalizability theory)に基づいた一般化可能性係数(generalizability coefficient; 以下G係数とする)(基準値は0.80以上)を用いて、評定者による4項目のルーブリック評価がどの程度安定的であるかを確認した。分散成分モデルによる推定にあたっては、正規性を仮定としない最小ノルム2次不偏推定量(MINQUE)を用いて、従属変数は評価者の評点を、変数因子は学生(p)、評定者(r)、評価項目(i)を投入して実施した。また、どの項目が評定に強く影響するかを確認するために、それぞれの分散成分の推定値が、分散成分全体に占める分散割合(%)を示した。

全体に対する分散割合の学生の主効果の「学生(p)」は、学生の得点のバラツキを示し、数値が高いほど学生間の個人差をとらえていることを、評定者の主効果の「評定者(r)」は評定者間においてバラツキがあるかを、評価項目の主効果の「評価項目(i)」は評価項目間でバラツキがあるかを意味する。また、学生と評定者の交互作用である「学生*評定者(pr)」は、評定者によって学生の順位が入れ替わる程度を、「学生*評価項目(pi)」は、学生の得手不得手などが影響して評価項目によって学生の順位が入れ替わる程度を、「評定者*評価項目(ri)」は、それぞれの評定者の評価項目に対する重要度の認識の違いの程度を意味する。すべての要因の交互作用と測定誤差を含む「学生*評定者*評価項目(pri)」は、その値が大きいくほど主効果や交互作用を説明できない部分が多くなり、G係数値に悪影響を及ぼすことを意味する。

なお、G係数は、下記の計算式で求めた(Nr; 評定者数3名、Ni; 評価項目数4項目)(池田 1994, リン 他 1992, 松下 他 2013, 日本アイ・ビー・エム株式会社 2016, 宇佐見 2011)。

$$G = \frac{p}{p + \frac{pr}{Nr} + \frac{pi}{Ni} + \frac{pri}{NrNi}}$$

(3) ルーブリック TEAM-P2016v 試作版の信頼性

ルーブリック TEAM-P2016v 試作版の4変数について合成変数として集約可能であるかを確認するために、評定者3名の評点の平均値を求めて主成分分析を行った。今回は正規性が確認されなかったため、数値変数間の非線形関係を前提とするカテゴリー主成分分析(categorical principal components analysis)を行った。成分負荷量は0.70以上を関係ありとし、説明された分散(固有値)の累積割合の基準値は70%以上とした。内的整合性は、固有値合計に基づいてカテゴリー主成分分析によって算出されたCronbachの α 信頼係数(以下、 α 信頼性係数とする; 基準値は0.80以上)と、元の変数の相関行列で確認した。

(4) ルーブリック TEAM-P2016v 試作版の妥当性

ルーブリック TEAM-P2016v 試作版の基準関連妥当性は、学生の TEAM-P2016v の自己採点を用いて、協同学習の経験頻度得点とチームワーク能

力尺度の因子別平均値との相関係数（スピアマンの相関係数）で確認した。

前述した通り（前述調査内容5.5.2(4)参照）、本来の妥当性検証では、評定者が評価したルーブリックTEAM-P2016v試作版と既存のルーブリックの観察得点の相関関係を確認するべきであるが、隣接概念を評価するルーブリックが存在しないために、学生が自己採点したルーブリックTEAM-P2016v試作版と『チームワーク能力尺度』得点を妥当性検証の参考値として用いた。

(5) 評定者評価と学生自己評価のルーブリック

TEAM-P2016v 試作版の得点比較

協同学習における学生のパフォーマンス評価のあり方への示唆を得るために、評定者3名の評点の平均値と学生の自己採点の相違を、マン・ホイットニーのU検定を用いて比較した。

以上の統計分析はIBM SPSS 24 (Statistics Base, Advanced Statistics, Categories, Exact Tests)を用いて行い、有意水準は5%とした。

6. 結果

図2は、対象の協同学習の経験頻度得点の分布を示した。中央値は9点で、中央値以上の度数は41名（59.4%）、0点から3点の度数合計は4名（5.8%）だった。

評定者が学生に対して実施したルーブリックTEAM-P2016v試作版の評点の一致率と評定者間比較および級内相関係数（ICC）は、表2に示した。全項目とも3名一致はおおよそ5割以上で、【②エンパワリング】と【③リフレーミング】は7割以上だった。評定者間の評点で有意差がみられたのは【①話す・聴く】と【④コンフリクト対応】で、両項目とも評定者Zが有意に低値だった。評定者3名のICCは0.792 ($p < 0.0001$)であった。なお、ある学生に対して評定者X・Yよりもワンランク低い評点を付けた評定者Zの【①話す・聴く】の備考欄には「積極的に話している時もあるが、他の学生が話している時にスマホをみていた」が、【④コンフリクト対応】の備考欄には「基本的に人の話しをあまり聴いていないので、グループワークが

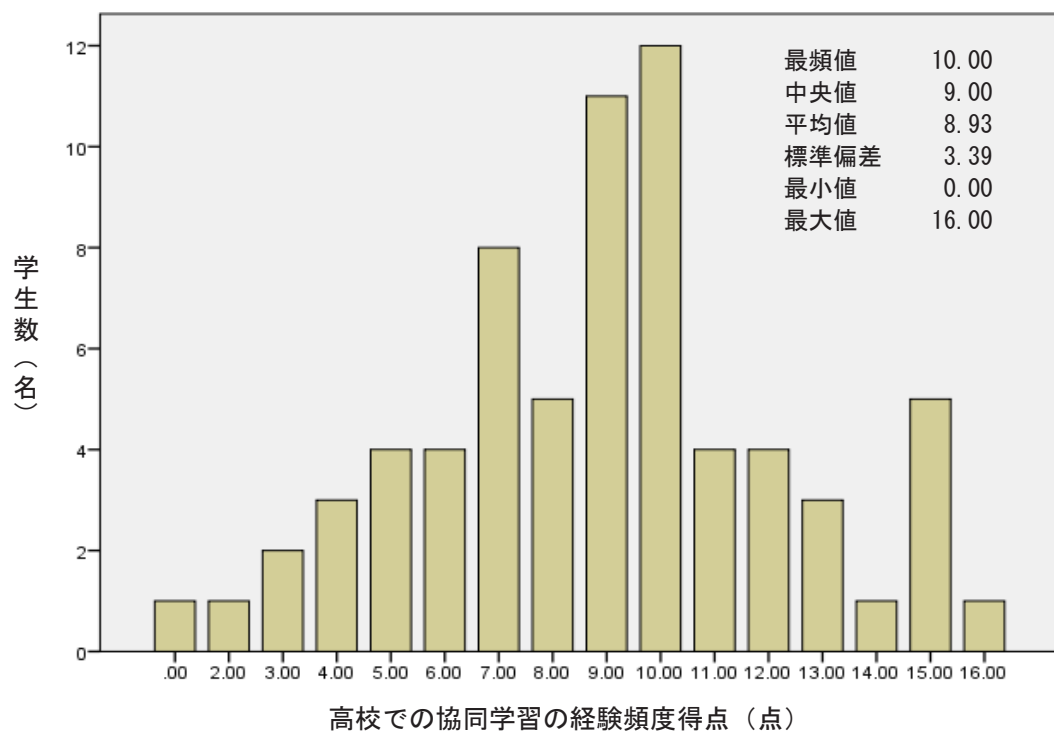


図2. 対象者69名の高校での協同学習[†]の経験頻度得点の分布

[†] 協同学習の代表的な6形態（①インフォーマル・グループ・ディスカッション、②シンク・ペア・シェア、③学生同士の教え合い技法、④グループ探求法、⑤ストラクチャード・プロブレム・ソルビング、⑥ディベート）の高校での経験頻度を点数化して（3点、頻回に経験した；2点、時々経験した；1点、数回位は経験した；0点、一度も経験なし）、それらを合算した合計点（0-18点）をヒストグラムに示した。

表2. 評定者3名が学生69名に対して実施したルーブリックTEAM-P2016v試作版の評点の一致率と評定者間比較および級内相関係数(ICC)

評定者間の一致率			評定者間比較					
一致度	度数	(%)	評定者ID	平均値	標準偏差	平均ランク	多重比較 [‡]	p値 [†]
【①話す・聴く】 Speaking Up & Active Listening								
全員不一致	0	(0.0%)	評定者X	2.67	0.505	2.09	*]**	p<0.0001
2名が一致	34	(49.3%)	評定者Y	2.78	0.449	2.26		
3名が一致	35	(50.7%)	評定者Z	2.38	0.488	1.65		
【②エンパワリング】 Empowering Team								
全員不一致	0	(0.0%)	評定者X	2.77	0.425	1.99	-	0.882
2名が一致	9	(13.0%)	評定者Y	2.77	0.489	1.99		
3名が一致	60	(87.0%)	評定者Z	2.78	0.415	2.01		
【③リフレーミング】 Reframing & Collaboration								
全員不一致	0	(0.0%)	評定者X	2.83	0.382	2.12	-	0.180
2名が一致	19	(27.5%)	評定者Y	2.74	0.533	1.99		
3名が一致	50	(72.5%)	評定者Z	2.68	0.469	1.90		
【④コンフリクト対応】 Coping with Conflict								
全員不一致	0	(0.0%)	評定者X	2.72	0.450	2.04	*]	p<0.0001
2名が一致	29	(42.0%)	評定者Y	2.83	0.419	2.20		
3名が一致	40	(58.0%)	評定者Z	2.54	0.502	1.76		
級内相関係数 (ICC) [§]			0.792	95%信頼区間		0.684 - 0.868	有意確率	p<0.0001

[†] 対応サンプルによるFriedmanの順位付けによる比較

[‡] 順位付けによるFriedmanの二元配置分散分析 *p<0.05, **p<0.01

[§] 評定者の平均測定値のICC (intraclass correlation coefficients) の基準値は0.70以上とした。

表3. ルーブリックTEAM-P2016v試作版の各分散成分の推定値と一般化可能性係数(学生数69名)

成分	分散成分の推定	全体に対する分散割合
学生 (p)	0.0875	37.8%
評定者 (r)	0.0069	3.0%
評価項目 (i)	0.0013	0.5%
学生*評定者 (pr)	0.0120	5.2%
学生*評価項目 (pi)	0.0342	14.7%
評定者*評価項目 (ri)	0.0094	4.1%
学生*評定者*評価項目 (pri)	0.0803	34.7%
計	0.2316	100.0%

一般可能性係数 (G係数) [†] 0.8197

[†] 分散成分分析の最小ノルム 2 次不偏推定量 (MINQUE) による推定方法を用いて, G係数 (Generalizability coefficient) の基準値は0.80以上とした. G係数は, 下記の計算式で求めた.

Nr (評定者数; 3名), Ni (評定項目数; 4項目)

$$G \text{ 係数} = \frac{p}{p + \frac{pr}{Nr} + \frac{pi}{Ni} + \frac{pri}{NrNi}}$$

進まないことに対して根気強く対応しているとは言いがたい」との記載があった。

表3に示すごとく、ルーブリックTEAM-P2016v試作版のG係数は0.8197であった。全体に対する分散割合は、「学生(p)」37.8%、「学生*評定者*評価項目(pri)」34.7%、「学生*評価項目(pi)」14.7%の順で高く、「学生*評定者(pr)」と「評

定者(r)」は5%未満であった。

表4に示すごとく、評定者3名のルーブリックTEAM-P2016v試作版の評点(平均値)によるカテゴリ主成分分析の結果では、第1主成分において4項目とも成分負荷量が基準値の0.70以上で、固有値%は73.9%、α信頼性は0.882であった。表5は、カテゴリ主成分分析によって算出され

表4. 評定者3名のルーブリック TEAM-P2016v 試作版の評点(平均値)によるカテゴリー主成分分析の結果(学生数69名)

	第1主成分の成分負荷量 [†]
【①話す・聴く】	0.9344
【②エンパワリング】	0.7397
【③リフレーミング】	0.8600
【④コンフリクト対応】	0.8918
固有値	2.955
固有値%	73.9%
Cronbachの α 信頼係数	0.882

カテゴリー主成分分析 (categorical principal components analysis)

[†] 基準値は、成分負荷量は0.70以上を、固有値%は70%以上、Cronbachの α 信頼係数は0.80以上とした。

表5. 評定者3名のルーブリック TEAM-P2016v 試作版の評点(平均値)によるカテゴリー主成分分析で算出された元の変数の相関行列(学生数69名)

	【①話す・聴く】	【②エンパワリング】	【③リフレーミング】	【④コンフリクト対応】
【①話す・聴く】 Speaking Up & Active Listening	1.000	0.518	0.840	0.809
【②エンパワリング】 Empowering Team		1.000	0.473	0.623
【③リフレーミング】 Reframing & Collaboration			1.000	0.613
【④コンフリクト対応】 Coping with Conflict				1.000
次元	1	2	3	4
固有値	2.955	0.612	0.340	0.094

カテゴリー主成分分析 (categorical principal components analysis)

た元の変数の相関行列の係数で、有意係数の範囲は0.473から0.840であった。

表6の基準関連妥当性の検討では、ルーブリック TEAM-P2016v 試作版の学生の自己採点と高校での協同学習の経験、チームワーク能力尺度得点の関係を参考値として採用した。その結果、高校での協同学習の経験と【③リフレーミング】のみで弱い正の相関関係が確認され ($r = 0.240$)、学生が回答したチームワーク能力尺度との間では「弱～中等度」の相関係数 ($r = 0.241-0.413$) が確認された。

表7に示すごとく、評定者3名の評点(平均値)と学生の自己採点を比較した結果、全項目で学生自己評価点が評定者よりも有意に低かった。

7. 考察

7.1 対象の背景

高校での協同学習の経験頻度得点では、約6割が中央値の9点以上で、ほとんど経験がないと考えられる0点から3点の学生は1割未満だった。従って、対象の9割近くが高校でなんらかの協同学習を経験していることになるが、その背景として、学習指導要領(文部科学省2008, 2009, 2013)の改訂が多分に影響していると考えられる。

本調査の対象となった現役学生は平成8年生まれで、中学入学の平成21年度から、いわゆる「脱ゆとり教育」が移行措置として始まり、高校入学の平成24年度には「高等学校等の新学習指導要領」にもとづいて、数学および理科科目の脱ゆとり教育が実施されている(完全実施は平成25年度

表6. ルーブリックTEAM-P2016v試作版の基準関連妥当性の検討: 学生69名の自己採点に対する協同学習の経験頻度得点とチームワーク能力尺度の因子別平均値との相関係数

	学生の自己採点によるルーブリックTEAM-P2016v試作版				合計
	【①話す・聴く】	【②エンパワリング】	【③リフレーミング】	【④コンフリクト対応】	
	Speaking Up & Active Listening	Empowering Team	Reframing & Collaboration	Coping with Conflict	
高校での協同学習の経験 [†]					
経験頻度得点 (0-18点)	0.221 p=0.067	0.054 p=0.658	0.240 p=0.046	0.148 p=0.223	0.212 p=0.081
チームワーク能力尺度の因子別平均値					
因子1 コミュニケーション能力	0.117 p=0.336	0.193 p=0.111	0.235 p=0.051	0.267 p=0.026	0.287 p=0.016
因子2 チーム志向能力	-0.072 p=0.559	0.270 p=0.0250	0.112 p=0.357	0.090 p=0.461	0.122 p=0.318
因子3 バックアップ能力	0.247 p=0.041	0.217 p=0.073	0.386 p=0.001	0.211 p=0.082	0.320 p=0.007
因子4 モニタリング	0.241 p=0.045	0.246 p=0.042	0.288 p=0.016	0.326 p=0.006	0.350 p=0.003
因子5 リーダーシップ能力	0.386 p=0.001	0.265 p=0.027	0.380 p=0.001	0.216 p=0.074	0.378 p=0.001
平均値の合計	0.291 p=0.015	0.290 p=0.015	0.413 p<0.0001	0.292 p=0.014	0.400 p<0.0001

スピアマンの順位相関係数

† 協同学習の代表的な6形態 (①インフォーマル・グループ・ディスカッション, ②シンク・ペア・シェア, ③学生同士の教え合い技法, ④グループ探求法, ⑤ストラクチャード・プロブレ・ソルビング, ⑥ディベート) の高校での経験頻度を点数化して (3点, 頻回に経験した; 2点, 時々経験した; 1点, 数回に経験した; 0点, 一度も経験なし), それらを合算した合計点 (0-18点) を用いた。

表7. ルーブリックTEAM-P2016v試作版の評定者3名の評点(平均値)と学生96名の自己採点の比較

	評点		p値
	評定者3名の平均値	学生の自己採点	
【①話す・聴く】 Speaking Up & Active Listening			
平均ランク	81.59	57.41	
中央値	2.67	2.00	p<0.0001
平均値	2.61	2.20	
最小値-最大値	1.67-3.00	1.00-3.00	
【②エンパワリング】 Empowering Team			
平均ランク	85.87	53.13	
中央値	3.00	2.00	p<0.0001
平均値	2.77	2.25	
最小値-最大値	1.67-3.00	1.00-3.00	
【③リフレーミング】 Reframing & Collaboration			
平均ランク	81.14	57.86	
中央値	3.00	2.00	p<0.0001
平均値	2.75	2.36	
最小値-最大値	1.67-3.00	1.00-3.00	
【④コンフリクト対応】 Coping with Conflict			
平均ランク	76.80	62.20	
中央値	2.67	2.00	p=0.022
平均値	2.70	2.36	
最小値-最大値	1.67-3.00	1.00-3.00	
合計点			
平均ランク	87.10	51.90	
中央値	11.33	9.00	p<0.0001
平均値	10.83	9.17	
最小値-最大値	7.00-12.00	4.00-12.00	

マン・ホイットニーのU検定

入学生から)。新学習指導要領では、従来の「ゆとり教育」でも「詰め込み教育」でもない「生きる力をはぐくむ教育」が方針として盛り込まれたことか

ら、主体的学習態度の養成や課題解決力の基盤となる思考力・判断力・表現力などを育成するための教育方法の一つとして導入されていたものと推

測される。

7.2 ルーブリック TEAM-P2016v 試作版の評定者間信頼性

ルーブリック TEAM-P2016v 試作版における評定者3名の ICC は 0.792、G 係数は 0.8197 で、いずれも基準値を満たしていた。

評定者間信頼性は、学生の主効果である「学生 (p)」が、すべての要因の交互作用と測定誤差を含む「学生 * 評定者 * 評価項目 (pri)」よりも数パーセントではあるが割合が高く、なおかつ、学生と評価項目の交互作用を示す「学生 * 評価項目 (pi)」よりも高値であったことが、基準値を満たす ICC が得られた要因と考えられた。また「評定者 (r)」と「学生 * 評価者 (pr)」が 5% 未満であったことは、評定者間でのバラツキや、評定者によって学生の順位が入れ替わる程度が低いことを意味することから、基準値以上の G 係数が得られたものと考えられた。

学生の記述課題（小論文、レポート）を評価した先行研究での G 係数は 0.62 から 0.65（松下 他 2013, 宇佐見 2011）、学生の主効果「学生 (p)」は 13.8%（松下 他 2013）と報告されており、教育経験が異なる3名の評定者による観察法で、学生のチーミング・プロセスを評価したルーブリック TEAM-P2016v 試作版の G 係数が基準値の 0.80 以上で、学生の主効果の分散成分割合が最も大きかったことは、評定者間信頼性を確保できているものと判断できた。

ICC と G 係数は基準値以上であったが、【①話す・聴く】と【④コンフリクト対応】の3名一致率が 50% 代で、評定者間比較でも有意差がみられたことから、この2項目の一致率が高くなれば、より評定者間信頼係数も高値になると考えられた。特に、Edmondson (Edmondson 2012, エドモンドソン 2014) がチーミング概念において基本に位置づけている「率直に意見を言うこと (speaking up)」と「積極的に意見を聴く (active listening)」の内容に該当する【①話す・聴く】は、学習中の状況場面が多いため、評定者 Z のコメントに記されていたように、観察した場面によって評点が影響することが考えられる。そのような影響を最小限にするために、協同学習に参加しているすべての学生の様子が満遍なく観察できる環境を整えること

と (机の配置など)、評定者はある特定のグループの様子のみを観察しないで、協同学習の時間経過の中で (例として前半・後半・終了直前など)、均等に観察していく必要があると考える。

一方、【④コンフリクト対応】は、グループ内の意見の対立などがなく、スムーズに進められている場合には観察しにくいと考えられることから、評定者間の3名一致率が比較的低下したと推測された。しかし評定者 Z のコメントに記されていたように、グループメンバーの発言や質問を聴くことがおろそかになっている学生は、コンフリクトの有無に関わらず、ワークの進捗に積極的に参加していないことを意味することとなる。これはカテゴリー主成分分析の元の変数の相関行列係数においても、【①話す・聴く】と【④コンフリクト対応】で高い相関係数が確認できたことから、チーミング概念の基本ともいえる【①話す・聴く】のパフォーマンスが高い学生は、【④コンフリクト対応】も高くなる傾向が考えられ、【①話す・聴く】の基本的態度を学生に習得させることが、チーミング力を向上させる可能性を示唆していると考えられた。

7.3 ルーブリック TEAM-P2016v 試作版の信頼性

ルーブリック TEAM-P2016v 試作版のカテゴリー主成分分析では、第1主成分において全項目の成分負荷量が 0.70 以上、固有値 %70% 以上、 α 信頼性係数は 0.8 以上、さらに元の変数間の相関行列の係数も中等度以上であったことから、チーミングを観察する説明率の信頼性は確保できたと考えられた。これは、チーミングの概念理論を基盤として、大学生の協同学習場面に共通する評価項目に焦点を当てたことが、項目間の相関関係を高めたものと考えられ、4つの【項目】の合計点で教育効果の推移を比較していくことも可能である。アクティブ・ラーニングの普及により、様々なメソッドの協同学習が実施されていくことが推測されるが、今後は、このような変化に対応したルーブリックの更なるヴァージョンアップが必要になると考える。

7.4 ルーブリック TEAM-P2016v 試作版の妥当性

研究方法で述べたように、隣接概念を評価する既存のルーブリックが確認できなかったために、基準関連妥当性の参考値として、ルーブリック

クTEAM-P2016v試作版の学生の自己採点と高校での協同学習の経験頻度、学生が回答したチームワーク能力尺度得点を用いて行った。その結果、高校での協同学習の経験得点は【③リフレーミング】との間に弱い相関関係が、『チームワーク能力尺度』は相関係数0.241から0.413の「弱～中等度」の相関関係が確認された。

高校での協同学習経験がルーブリック得点の予測変数として十分でなかった要因として、協同学習経験の「種類」と「頻度」の量的側面のみを質問して、協同学習経験をどのように、どの位効果がでるようになってきたかの質的側面を測定していないことが影響していると考えられる。高校での「探求学習活動」経験と大学初年次の満足度や充実感との関係をパス解析で明らかにした木村(2010)の調査でも、高校での「探求学習活動」経験が4変数からなる『大学満足』(サービス満足、友人満足、設備満足、授業満足)に与える直接効果のほかに、「積極発言」による間接効果が確認された結果から、高校で「探求学習活動」を経験したからといって、単線的に大学入学後の充実が保障されるというわけではなく、「探求学習活動」を通して「積極発言」を行うような自律的態度を涵養していくことの重要性を示唆しており、本研究の結果を裏付けるものとなっている。

もう1つの外的基準である『チームワーク能力尺度』については、ルーブリックの4項目と合計点との間に「弱～中等度」の相関関係がいくつか確認されたことから、ある程度の予測力を有するものと考えられた。今回使用した『チームワーク能力尺度』は、個人がチームに属したときに発揮する能力(コンピテンシー)を測定するもので、既存の尺度にない総合的能力(コミュニケーション能力、チーム志向能力、バックアップ能力、モニタリング能力、リーダーシップ能力)が構成されていることが、ある程度の相関関係が得られた要因と考えられた。ただし、強い相関関係が得られなかったのは、回答者が特定のチームに所属してなくても日常的な個人のチームワーク能力を点数化できる本尺度の汎用性の高さが、ルーブリックによる協同学習プロセスに対するセルフチェック得点との差異につながった可能性が考えられる。

なお、我が国で公開された既存のルーブリックにおいても妥当性を検証しているものはない

(松下 他 2013, 宇佐見 2011)。海外の学術文献でも、内容的妥当性や、対象の属性によるルーブリックの得点比較にとどまっておらず(Curran et al 2011, Yeung et al 2016)、外的基準の設定による基準関連妥当性の検証を今後の研究課題として述べている。また、Lasater(2007)が開発した看護の臨床判断力を評価する11項目の英語版ルーブリック(Lasater Clinical Judgment Rubric; LCJR)の韓国語バージョン(K-LCJR)を考案したShinら(2015)も、妥当性を検証したルーブリックがないことを指摘しつつ、本来、評定者の観察用に作成されたK-LCJRを、学生152名の自己採点によって構成概念妥当性を確認している。以上のことは、ルーブリックの開発とその統計学的検証が、端緒についた分野であることを反映しており、今後、質の高いルーブリックが増えていくことによって、基準関連妥当性の検証が進むものと思われる。

7.5 評定者と学生自己評価得点との比較からみた教育評価への示唆

協同学習における学生のパフォーマンス評価のあり方への示唆を得るために、評定者評価と学生の自己評価点を比較した結果、全項目で学生自己評価点が評定者の評価点よりも有意に低かった。

本研究において、学生自己評価が評定者評価よりも低かった背景として、協同作業に対する学生の苦手意識や有意味感の低さなど、行動面では把握しにくい学生の認識が関係している可能性が考えられる。看護学生131名の協同作業に対する認識の影響要因を明らかにした調査でも(會田 他 2017)、大学での学習満足度やソーシャルサポート、対人葛藤方略スタイルおよび仮想的有能感の他者軽視などが有意変数として確認され、協同作業に対する認識には、学生の職業アイデンティティやコミュニケーション力、過去のネガティブ体験などの多様な要因が影響している可能性が報告されている。このことから、評定者が学生のパフォーマンスを適切に評価し、フィードバックしていくことは、学生の協同作業に対する苦手意識の克服や自信につながっていく可能性が考えられ、ルーブリックTEAM-P2016v試作版を利用していく意義を示唆していると考えられる。

医療保育科短期大学の女子学生48名が履修す

るオペレッタ授業の学生自己評価と教員評価の相違を明らかにした青井ら(2013)の調査では、本研究とは反対に、12項目中9項目において学生の自己評価点が教員評価よりも有意に高値であった。この要因として、青井らは、学生からの質問・相談・不満などに対応した教員の学生に対する印象が、学生の低評価に影響している可能性や、学生が自身の気持ちに折り合いをつけながら問題解決している実態を、教員が把握しきれていない可能性などを考察している。

このような教員(評定者)の主観的感覚や個別的状况などによる影響を可能な限り緩和する目的としてルーブリックTEAM-P2016v試作版を開発した。しかし、本研究の評定者Zの「他の学生が話している時にスマートフォンをみている場面」に対するとらえ方が示すように、同じ評価基準を使用しても評点において多少の違いが生じる可能性は避けられない。その意味で、協同学習中の学生のパフォーマンスを評価する場合は、共通した評価基準を活用することと、協同学習の様子を適切に観察していくことの重要性が改めて確認できたといえる。

さらに、ルーブリックによる学生の自己評価を取り入れていくことで、教員評価との乖離の程度を把握できる。このことによって、フィードバックの際に教員と学生との話し合いが可能となり、教員が期待している内容をより具体的に学生が認識できることと、学生のみでなく、教員自身の教育内容や質の振り返りにもつながるメリットがあると考えられる。

7.6 ルーブリックTEAM-P2016v試作版の考案の意義と今後の課題

以上の考察を総合的に踏まえて、ルーブリックTEAM-P2016v試作版は、基準関連妥当性において課題は残されたものの、協同学習中の大学生のチーミング・プロセスを評価するパフォーマンス評価指標として適用可能であると考えられた。評価する【項目】を4項目に限定して、できている項目の数で評価していくので、観察する所要時間にもよるが、複数の教員でなくとも、一人の教員での利用も可能と考えられ、教育現場の実状にあった利用が可能と思われる。

特に協同学習におけるパフォーマンス評価にお

いては、学生の自己評価のみを主軸とすると過小評価に偏る可能性が考えられ、教員間での差異は避けられないものの、教員が期待する教育評価の視点と学生の自己評価とのギャップを小さくする意義はあると考える。また、協同学習における学生の課題が、Edmondson (Edmondson 2012, エドモンドソン 2014) がチーミングの重要な基礎として重要視している【①話す・聴く】(率直に意見を言い積極的に聴く; Speaking Up & Active Listening)を強化することによって高められる可能性が示唆されたことから、チームワーク力を評価する学位授与方針のルーブリックを新たに開発し発展させていくことによって、社会から期待される大学教育の質を高めていく一助になると考える。

今回は、対象に対する教育上の倫理的配慮やフィールド確保の困難などから、協同学習に参加するB大学の看護学生2年生を対象として検討した。ある特定大学の特定分野に所属し、特定の協同学習課題に参加している学生を対象としたことが、どの程度のセレクションバイアスとして影響を及ぼしているのか、他の医療・学問領域の学生や多職種連携教育などのフィールドでの検証が必要である。また、評定者数は評価の実行可能性を低下させない最少人数の3名(松下他2013)で、なおかつ、カリキュラム上の教育課題や学生の学習プロセスに悪影響を及ぼさないよう観察を行ったが、一度に観察する学生数や観察方法などによっても影響を受けるのかについても統計的に検討していく必要が考えられた。

告知

本研究は、信州大学医学部医倫理委員会で承認を得ている〔承認番号3389〕。本研究に関して、著者らは開示すべき利益相反はない。

謝辞

本調査にご協力を頂きました学生、関係者の皆さまに深謝いたします。なお、本研究はJSPS科研費JP25463296の助成を受けたものです。

引用文献

阿部恵子, 若林英樹, 西城卓也 他(2012). Trait Emotional Intelligence Que-SFと Jefferson Scale of Physician Empathyの日本語版開発と信頼性・妥当性の検討. 医学教育43, 351-359. DOI: 10.11307/mededjapan.43.351

Agency for Healthcare Research and Quality (AHRQ) (2014). TeamSTEPPS® Instructor Manual: Team Performance Observation Tool. <https://www.ahrq.gov/teamstepps/instructor/reference/tmpot.html> (accessed: 2016-2-28).

相川充, 高本真寛, 杉森伸吉 他(2012). 個人のチームワーク能力を測定する尺度の開発と妥当性の検討. 社会心理学研究27, 139-150. DOI: 10.14966/jssp.KJ00008019341

會田信子, 三好沙知, 河地美紀 他(2017). A大学看護学生の協同学習に対する認識と影響要因. 医学教育48, 59-69.

American Association of Colleges & Universities (n. d.). VALUE (Valid Assessment of Learning in Undergraduate Education). http://www.in.gov/che/files/All_VALUE_Rubrics.pdf (accessed: 2016-2-28).

青井則子, 入江慶太, 秋政邦江 他(2013). 総合表現(オペレッタ)における授業開発-学生の自己評価と教員評価の差異の検討-. 川崎医療短期大学紀要33, 55-60. DOI: info:doi/10.18928/00000236

エリザベス・F. バークレイ, パトリシア・P. クロス, クレア・H. メジャー (2009). 安永悟(監訳). 協同学習の技法-大学教育の手引き-, pp3-20. ナカニシヤ出版, 京都.

中央教育審議会(2008). 学士課程教育の構築に向けて. http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/houkoku/080410.htm (最終閲覧日: 2016年2月28日)

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～. http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/

toushin/1325047.htm (最終閲覧日: 2016年2月28日)

Curran V, Hollett A, Casimiro LM et al (2011). Development and validation of the interprofessional collaborator assessment rubric (ICAR). J Interprof Care 25, 339-344. DOI: 10.3109/13561820.2011.589542

Edmondson AC (2012). Teaming: How Organizations Learn, Innovate, and Compete in the Knowledge Economy, pp11-43. Jossey-Bass, San Francisco.

エイミー・C. エドモンドソン (2014). 野津智子(訳). チームが機能するとはどういうことか-「学習力」と「実行力」を高める実践アプローチ, pp21-108. 英治出版, 東京.

Frankel A, Gardner R, Maynard L et al (2007). Using the Communication and Teamwork Skills (CATS) Assessment to measure health care team performance. Jt Comm J Qual Patient Saf 33, 549-58. DOI: [http://dx.doi.org/10.1016/S1553-7250\(07\)33059-6](http://dx.doi.org/10.1016/S1553-7250(07)33059-6)

藤野ユリ子(2005). 看護学生がグループワークで感じる困難と満足との関係. 日本看護学教育学会誌15, 1-14.

Guise JM, Deering SH, Kanki BG et al (2008). Validation of a tool to measure and promote clinical teamwork. Simul Healthc 3, 217-223. DOI: 10.1097/SIH.0b013e31816fdd0a

芳我ちより, 舟島なをみ(2007). 学生間討議を中心としたグループ学習における教授活動の解明-看護基礎教育において展開される授業に焦点を当てて. 看護教育学研究16, 15-28. DOI: http://doi.org/10.19015/jasne.16.1_15

市川茂子, 嶋松陽子, 和田奈津子(1994). 主体性を引き出すグループ学習指導. 看護教育35, 141-145. DOI: <https://doi.org/10.11477/mf.1663900788>

飯岡由紀子, 亀井智子, 宇都宮明美(2016). チームアプローチ評価尺度(TAAS)の開発-尺度開発初期段階における信頼性と妥当性の検討-. 聖路加

- 看護学会誌19(2), 21-28.
- 池田央(1994). 現代テスト理論, pp28-34. 朝倉書店, 東京.
- 池田浩, 古川久敬(2009). 組織における「チーム力」: 課題の変化と成果を意識したチーム能力の概念化と測定尺度の開発. 産業・組織心理学会第25回発表論文集, 139-142.
- デイヴィット・W. ジョンソン, ロジャー・T. ジョンソン, ジョンソン・E. ホルベック(2010). 石田裕久, 梅原巳代子(訳). 学習の輪 -学び合いの協同教育入門- 改訂新版, pp7-28. 二瓶社, 大阪.
- 木村拓也(2010). 高校時代の探求活動経験が初年次学生に与える影響 -JFS2008の結果から. *Journal of Quality Education* 3, 77-93.
- 小味慶子, 大西麻未, 菅田勝也(2011). 医師と看護師の協働に対する態度: Jefferson Scale of Attitudes toward Physician -Nurse Collaboration 日本語版の開発と測定. *医学教育* 42, 9-17. DOI: 10.11307/mededjapan.42.9.
- 小山真里子(1995). 効果的なグループ学習を促進するための教師の関わり. *Quality Nursing* 1, 23-27.
- 國澤尚子, 村本淳子, 澤井映美 他(1996). 学生のグループ学習能力に対する1年間の認識の変化. *東女医大看護研究紀要* 18, 1-6.
- Lasater K (2007). Clinical judgment development: Using simulation to create an assessment rubric. *J Nurs Educ* 46, 496-503.
- ロバート・L. リン, サミュエル・メシック, リー・J. クロンバック(1992). 池田央, 柳井晴夫, 藤田恵壘 他(訳). 教育測定学 原著第3版 上巻, pp182-200. 学習評価研究所, 東京.
- Malec JF, Torsher LC, Dunn WF et al (2007). The mayo high performance teamwork scale: reliability and validity for evaluating key crew resource management skills. *Simul Healthc* 2, 4-10. DOI: 10.1097/SIH.0b013e31802b68ee
- 松下佳代(2007). パフォーマンス評価 -子どもの思考と表現を評価する-, pp6-14. 日本標準, 東京.
- 松下佳代(2012). パフォーマンス評価による学習の質の評価 -学習評価の構図の分析にもとづいて-. *京都大学高等教育研究* 18, 75-114.
- 松下佳代, 小野和宏, 高橋雄介(2013). レポート評価におけるルーブリックの開発とその信頼性の検討. *大学教育学会誌* 35, 107-115.
- 松田麗子, 牧野典子(2012). 保健看護学科成人看護学実習のグループ活動における協同的な学びの効果. *中部大学教育研究* 12, 99-104.
- 三沢良, 佐相邦英, 山口裕幸(2009). 看護師チームのチームワーク測定尺度の作成. *社会心理学研究* 24, 219-232. DOI: 10.14966/jssp.KJ00005381247
- 文部科学省(2008). 平成20年3月現行学習指導要領・生きる力. http://www.mext.go.jp/a_menu/shotou/new-cs/idea/ (最終閲覧日: 2016年11月17日)
- 文部科学省(2009). 「学士課程教育の構築に向けて」中央教育審議会答申の概要. http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/siryo/attach/1247211.htm (最終閲覧日: 2016年11月17日)
- 文部科学省(2012). 大学教育部会(第11回)大学教育部会の審議のまとめについて(素案). http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/015/attach/1318247.htm (最終閲覧日: 2016年2月28日)
- 文部科学省(2013). 高等学校等の新学習指導要領の実施に当たって(通知). http://www.mext.go.jp/b_menu/hakusho/nc/1343618.htm (最終閲覧日: 2016年11月17日)
- 長濱文与, 安永悟, 関田一彦 他(2009). 協同作業認識尺度の開発. *教心理研* 57, 24-37. DOI: 10.5926/jjep.57.24
- 日本アイ・ビー・エム株式会社(2016). IBM SPSS Advanced Statistics 24, pp21-24. ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/24.0/ja/client/Manuals/IBM_SPSS_Advanced_Statistics.pdf

沖裕貴(2014). 大学におけるルーブリック評価導入の実際 -公平で客観的かつ厳格な成績評価を目指して-. 立命館高等教育研究14, 71-90.

関田一彦, 安永悟(2005). 協同学習の定義と関連用語の整理. 協同と教育1, 10-17.

Shin H, Park CG, Shim K (2015). The Korean version of the Lasater Clinical Judgment Rubric: A validation study. NURSE EDUC TODAY 35, 68-72. DOI: <http://dx.doi.org/10.1016/j.nedt.2014.06.009>

Sigalet E, Donnon T, Cheng A et al (2013). Development of a Team Performance Scale to Assess Undergraduate Health Professionals. Academic Medicine 88, 989-996. DOI: 10.1097/ACM.0b013e318294fd45

ダネル・D. スティーブンス, アントニア・J. レビ(2014). 佐藤浩章(監訳), 井上敏憲, 俣野秀典(訳). 大学教員のためのルーブリック評価入門, pp2-36. 玉川大学出版部, 東京.

太幡直也(2017). 大学生のチームワーク能力を向上させるトレーニングの有効性. 教育心理学研究 65, 305-314. DOI: 10.5926/jjep.64.118

田中耕治(2010). 新しい「評価のあり方」を拓く - 「目標に準拠した評価」のこれまでとこれから -, pp4-12. 日本標準, 東京.

The Council for Aid to Education (n. d.). CLA Scoring Criteria: Performance Task. <http://cae.org/images/uploads/pdf/CLAScoringCriteria.pdf> (accessed: 2016-2-28)

塚崎恵子(1996). 地域看護学におけるグループワーク学習の効果. 看護展望21, 1131-1136.

津村俊充(2012). プロセス・エデュケーション - 学びを支援するファシリテーションの理論と実際 -, pp9-18. 金子書房, 東京.

宇佐美慧(2011). 小論文評価データの統計解析 - 制限字数を考慮した測定論的課題の検討 -. 行動計量学38, 33-50. DOI: 10.2333/jbhmk.38.33

マイケル・A. ウェスト(2014). 下山晴彦(監訳),

高橋美保(訳). チームワークの心理学 -エビデンスに基づいた実践へのヒント-, pp26-28. 東京大学出版会, 東京.

山口裕幸(2007). チーム・コンピテンシーと個人のチームワーク能力. 教育テスト研究センター第1回研究会報告書, 1-14.

Yeung E, Kulasagarem K, Woods N et al (2016). Validity of a new assessment rubric for a short-answer test of clinical reasoning. BMC Med Educ 16, 192. DOI: 10.1186/s12909-016-0714-1



著者連絡先

〒390-8621

長野県松本市旭3-1-1

信州大学 学術研究院 医学保健学域 保健学系

會田 信子

aida@shinshu-u.ac.jp