

別紙 4

報告番号	※	第	号
------	---	---	---

主 論 文 の 要 旨

論文題目	クラスルームテストの作成・評価方法の検討： 多枝選択式項目作成ガイドラインと識別力指標について
氏 名	坪田 彩乃

論 文 内 容 の 要 旨

テストとは、教育成果の実態について目に見える形での情報を与えてくれる唯一の道具である(池田,2006)。この道具が適切でない場合、様々な問題が生じる。たとえば、受検者の能力を引き出すのに十分ではない試験問題であったとき、受検者の能力の実際が反映されない評価をつける可能性がある(池田, 1992)。効果的にテストを使用するためにも、テスト項目の作成から、テスト実施後の評価に至るまで、その手法について十分に検討される必要がある。

テストの種別のひとつに、学校現場で教師が作成するクラスルームテストがある。これは、クラスルーム内で行われるテストを指す。大学入学共通テストを始めとする大規模テストでは、テストの作成や実施、成績評価を複数の人間の手を通して行う一方で、クラスルームテストではその一連の流れを教師が行うという点が特徴である。そのため、教師ひとりひとりがテストの作成から分析についての知識やスキルを持っていることが望ましい。

テストの作成をする際には、測定したい能力以外の要因が正誤に影響しないような項目を作成し、テストが実施の意図に沿うものであったかを検討する必要がある。たとえば、項目に Flaw(欠点)があることで、テストが測定道具としての機能を十分に果たさないという指摘がある(Rush, Rankin, & White, 2016; Tarrant, Knierim, Hayes, & Ware, 2006)。また学校現場などでは、そもそもテストの評価はほとんどされていない。そこで本研究では、テストの専門家のみならず、テストの非専門家がテストの作成・実施・分析を行う際に有用な知見を示すことを目的として、「よいテスト」を作成するために、テスト項目の作成時ならびにテスト実施後の評価にかかわる問題について検討を行った。本論文は 5 つの章から成る。各章の内容は以下の通りである。

第 1 章では、学校現場で教師が作成するクラスルームテストについて、教育評価の観点から、テストが行われる目的、テストが児童・生徒に及ぼす影響について整理した上

で、テストにまつわる問題点について紹介した。クラスルームテストでは、授業過程の複数の段階で児童・生徒の学力を測定し、適切にフィードバックをするという役割を担う。しかし、クラスルームテストの作成は、教師個人の経験や勘に基づいて行われており、それが適切なテストであったのかの評価はほとんど行われていない。テストから得られた評価が実態を反映していないとき、児童・生徒の処遇に影響を及ぼすことがある。また、ピグマリオン効果(Rosenthal & Jacobson, 1968)など、学力そのものへ影響を及ぼすことがある。こうした実態を反映しない評価を作成・実施しないためにも、テスト作成時に生じる問題点や、テストの評価を行わないことで生じる問題点について述べた。

第2章では、「よいテスト」とはどのようなものであるのかを、テストの基本設計・目的・項目形式ならびにテスト実施後の評価の観点から検討した上で、よいテストを作るための方法について整理した。よいテストについて、日本テスト学会(2010)は、明確な目的の上で、その目的を果たすように作成・実施・採点・運用されているテストであるとし、小泉(2017)は、信頼性・妥当性・実用性を兼ね備えたものであるともいう。これらを兼ね備えたテストを作成するための方法のとして、テスト項目を作成する際に有用となる項目作成ガイドライン(Haladyna & Rodriguez, 2013)ならびに、テストが作成者の意図に沿っていたかを検討するための項目分析の必要性について論じた。

第3章では、研究1としてテストの作成に関する問題について検討をした。項目作成ガイドラインのうち、多枝選択式項目に関するものについて、ガイドラインに準拠していない項目であることが受検者に気が付かれることなく、項目正答率・識別力に影響するガイドラインを明らかにした。荒井(2015)は、ガイドラインでは用いない方が良いとされる項目形式であっても、場合によってはそれを使用することは妥当だと考える項目作成者がいたという。どのガイドラインがより項目特性へ影響を及ぼすかを明らかにすることで、重視すべきガイドラインが分かり、作題者の負担を減らしながらも、受験者の能力をより引き出せる項目作成に繋がる。

多枝選択式項目作成ガイドライン(Haladyna & Rodriguez, 2013)にある27のガイドラインのうち、25のガイドラインについて、同じ測定内容でありながらガイドラインに準拠／非準拠と項目からなる3パタンのテストを作成し、477名の受検者に行った。その結果、「各設問の内容は互いに独立であること」や、「『でない』『～以外』などの否定表現を用いないこと」といったガイドラインにおいて、非準拠の項目であるということが受検者にあまり気が付かれずに、項目正答率や識別力に影響を及ぼした。また、一部のガイドラインに非準拠である項目の中には、テストワイズネスといった、項目が本来測定したい能力とは異なる要因によって正答できている可能性が示唆された。

第4章では、研究2として、簡便に扱うことができる識別力指標(D 指標)につい

での検討を行った。項目の識別力指標はその項目がどの程度受検者の能力に基づいて正誤を判別できるかどうかというものであるが、教師にとってあまり馴染みがない。クラスルームテストでは非専門家がテストの評価を行うことを踏まえ、項目の識別力を評価する方法として簡便なアプローチが望ましい(Diederich, 1973)。このことから、テストの受検者について能力値を上位群・中位群・下位群と分けた際の、上位群と下位群の正答率についての差をとったものである D 指標が有用だと考えられる。 D 指標を算出するためには、受検者を 3 群に分ける必要があるが、限られた項目数で行われるテストでは、正答数が同じである同点の受検者が複数になることが予想される。そのため、同点の受検者をどのように扱うことが、より真値を反映した生徒の群わけの方法になるのか、また、それらを達成するために望ましいテストの条件について、 D 指標という側面から検討を行うことを目的とした。

まず、同点の受検者は全員同じ群となるよう群分けを行った場合の D 指標についてシミュレーションを用いて確認したところ、受検者数によらず一部の群に対象となる受検者が存在しなくなるなどの問題が見られた。次に、受検者集団を Kelly(1939)で示された比率となるように同点の受検者であっても異なる群とする群分けを行った場合の D 指標についてシミュレーションを用いて確認した。その結果、完全にランダムによる群分けをするよりも、項目の識別力に関する情報を用いた群分けをした方が真値に近い結果になることが明らかとなった。また、項目数が多いほど、かつ、項目の識別力が高いほど、真値に近い結果となることが明らかになった。そこで、実際のテストデータを用いて、シミュレーションと同様の結果となることを確認し、識別力を用いた群分けの有用性について確認した。

第 5 章では、以上の研究を踏まえ、テストの作成ならびに評価についての研究成果から、クラスルームテストを作成・実施するための方略について総合的に検討を行い、本研究の限界について述べた。クラスルームテストを行う上では、識別力の高い項目を多く用いることで項目分析に耐えるテストとなることから、多くの項目を作成する必要性が示唆された。項目作成ガイドラインを参照しながら項目を作成することで、よい項目を作成することに繋がる。また、一部のガイドラインに準拠しないことで、実際の項目がもつ機能以上の識別力となることがあることに留意する必要がある。なお、本研究で項目への影響が大きいとされたガイドラインであっても、科目によっては準拠することが難しいことも生じることから、個別の科目の特徴を踏まえたガイドラインの影響の大小について検討することの重要性について議論した。

