IEA COMPASS
教育概要
Researching education, improving learning
第 10 2020 年 9 月

国際大規模教育評価：
摘要ガイド

概要

国際大規模評価（ILSA）は、政策立案者とその他の教育ステークホルダーが、教育改革のためにエビデンスベースの意思決定を周知するための最も重要なツールの1つである。しかし、その重要性が認められ、 ILSAデータが広く用いられているにもかかわらず、評価結果が誤解または誤って解釈されることが時としてある。そこで、このILSA の摘要ガイドが用意された。このガイドは、調査内や調査間の重要な差異や共通性、評価にあたっての限界、そして、ILSAが教育評価と改革のための最も有意なツールの一部であり続けている理由を浮き彫りにする。また、国際教育到達度評価学会（IEA）の主な調査、アプローチ、組織的構造に焦点を当て、これらを他の ILSA と比較する。

国際教育到達度評価学会（IEA）
（本部：アムステルダム）
ウェブサイト：www.iea.nl
フォローする：

@iea_education
IEAResearchInEducation
IEA

要点

- IEA は、1958年に初めて ILSA 調査を実施したが、ILSAの課題は時代と共に変化し、新たなものに常に移り変わっていく。直近では、フォーカスがインプットではなくむしろ、教育成果に移ってきた。したがって ILSA の調査結果は、その時求められていたものが何かという文脈の中で理解すべきである。
- 組織と評価へのアプローチにはかなりの違いがある。レビュープロセスから意思決定や費用に至るまで、調査結果の理解にあたっては根本的な違いを考慮に入れるべきである。
- 主要な ILSA には、そのプロセスにおける各ステップを達成するために慎重に規定された高い品質基準を特徴とする共通の手法が見られ、さらにデータには、調査結果の解釈と報告を支援する目的で、データを支持する詳細な技術報告書が加えられる。
- 調査結果の理解と報告の際、ILSA の限界を認識する必要がある。とりわけ、教育制度の全体的な質を特定する尺度として ILSA の結果を考察する際には注意が必要である。
- 限界はあるが、大規模評価は、堅固な国際的な枠組みの中で経時的にシステムが監視され、おおむね、どの政治制度からも独立しており、データは自由に一般公開されるという点で特異な調査である。ILSAのデータを適切に分析し、理解すれば、それを反映させた政策決定や研究を行う貴重な機会を得ることになる。

序論

国際大規模教育評価は、世界中の教育能力を評価する実証的調査である。データは、教育に関する啓発を政策立案者、教育研究者、一般人に与える目的でさまざまな方法で用いられている。しかし、ILSA データが広範に使用されている一方で、調査結果が誤って解釈されたり、報告されたりすることがしばしば生じている。異なる調査による結果が同一のことを意味するものだと報告あるいは解釈されることが時としてあるが、調査結果の使用の際には重要な相違を検討する必要がある。国際大規模評価の分野のリーダーであるIEAが作成したこのガイドには、ILSA の調査結果と、その解釈と報告の方法をより良く理解するために考慮すべきさまざまな要素が説明されている。つまり ILSAとは何か、その発展の歴史、アプローチ面の違いと共通点、組織、手法、注意を向けるべき評価の限界などである。また、世界中の教育制度と学生たちの学習到達度を理解するため、また工ビデンスベースの変化を周知するためにILSA が重要かつ関連性の高いツールであると言える理由も説明される。

ILSA とは

ILSA は、特定学科における学生の学習到達度を評価し、学生についての付加的なデータを収集して調査結果の背景を提供する調査である。さらに、教師、校長や学長、教育システムについての詳細な関連情報を収集することもある。統計上有効な調査結果を提供するために、調査対象となる学校（通常、150～200校）がそれぞれの参加国または教育制度から抽出され、各学校から調査対象サンプルとなる学生が抽出される。いくつかのクラス全体がサンプルとして抽出される場合もあれば、各クラスから学生が抽出される場合もある。

ILSAの沿革

1958年、最初の ILSA である、IEA のパイロット12か国調査（ Pilot Twelve-Country Study ）（Foshay et al. 1962）が研究者グループによって実施された（Husén 1983; IEA 2018）。ハンブルク（ドイツ）の UNESCO 教育研究所（Institute for Education）にさまざまな分野の学者たちが一堂に会し、ある試験的な研究を実施することを決定した。その研究の目的は、多種多様な国や文化間の学習成果を比較することが可能かどうかを見定めることだった。その研究者たちは、さまざまな言語に翻訳するのが一番容易で、それゆえに各国同士の有効な比較が可能になると考え、数学における学生の学習到達度を評価することにした。研究者たちは、国際的な評価を通して何を学ぶことができるかを見いだすことを目的とし、同時に対象国が互いに学び合うことを期待した。Torsten Husén は、「概括的に言えば、こうした国際的な調査によって、教育専門家たち（広い意味で言えば、教育計画や教育政策決定の責任を負う関係者を含む）は他の国々の教育経験から恩恵を受けることができる。教育的成果に関連する変数の多くが、初めて標準化された方法で定量化されるので、教育専門家たちはより客観的に自身の教育制度を見ることができるようになる」と述べている（Husén 1967, pp. 13-14）。

最初のいくつかの ILSA は、限られたリソースが与えられた研究者によって実施され、教育調査に対する学術的な関心を満足させたが、1990年までには、教育関係の政策立案者たちが、ILSAには有益なエビデンスベースのデータを提供する潜在的な可能性があることを認識するようになった。それら政策立案者たちは、ILSA に対し金銭的サポートをすることで、早急な調査結果を期待した。初期の学術調査報告の中には、調査実施から最長8年経ってから公表されたものもあったが（Anderson et al. 1989）、こうして関心が広まることで、可能な限り早い調査結果の公表を求める新たな圧力が生まれた。こうした流れの中で、教育制度の動向に対する関心が生まれ、それがTIMSSを立ち上げるIEAの取り組みにつながった。TIMSSは1995年に初めて実施され、続いて4年後の1999年に行われた。1回目と2回目の間を4年とした理由は、TIMSS が1995年においてグレード4と8の学生を評価し、1999年にグレード8の学生のみを評価することで、グレード8の学生の2つの群の調査結果を比較し、またグレード4だった学生たちの4年後の進展を評価するためだった。

また1990年代、経済成長に対する理解を深めるために、しっかりした教育データを確保しようという世界的な関心が新たに生まれた。そして、有数の国際経済組織のである OECD は、教育および教育成果を測定することをいっそう重視するようになり、2000年に、自身の評価調査である PISA を立ち上げた。それ以前、OECDが教育研究を目的として使用したソースは、 IEA 調査や自身の年次発行物である Education at a Glance などにとどまっていた。しかしながら同組織は、学校が指導する内容を評価するのではなく、むしろ近代経済において効果的に働くのに必要なスキルにフォーカスすることに踏み切った。例えば、OECDは、15歳の学生が読解、数学および科学リテラシーにおいて知っておくべきと考える内容を定め、PISAによってそれに関する評価ができるとしている。この評価調査は3年ごとに実施されている。当初、調査は OECD 諸国を対象としていたが、今ではますます多くの非OECD 諸国が評価調査に参加するようになっている。

1990年代末、そして2000年代にかけ、中南米の LLECE（Laboratoria Latinoamericano de Evaluación de la Calidad de la Educatión）、PASEC（Programme d'Analyse des Systèmes Educatifs des Pays de la CONFEMEN）およびアフリカの SACMEQ（Southern African Consortium for the Measurement of Education Quality）を含め、地域レベルの ILSA も実施されるようになった。

最近のこととして、2015年に持続可能な開発目標（SDGs）が国際連合（UN） によって宣言されて以来、評価の重心が大きくシフトした。ミレニアム目標とは対照的に、SDGs は、教育への支出等のインプットではなくむしろ、教育制度における教育上の成果、つまりアウトプットを重視している。その結果、全ての国は、必要最低限の習熟度に達した学生がどれほどいるかを明らかにすることが強く期待されるようになり、アウトプット重視の傾向は明らかにILSA の新たな課題となっている。この点で IEA は、UNESCO 統計研究所（Institute for Statistics）、LLECE および PASEC と連携し、Rosetta Stone プロジェクト1 の実施を通して積極的な役割を果たしている。その目的は、地域の数学と読解評価による結果を TIMSS および PIRLS スケールで置き換える対応表を作成することである。

さまざまなアプローチ = さまざまな評価

最もよく知られている ILSA には数々の共通点が見られるが、同時に、大きな相違も存在している。さまざまな教育制度を念頭に置いて調査結果を比較する場合、それらの相違を考慮する必要がある。以下の表は、IEA および OECD によって実施される調査結果を比較する際、調査結果の利用者が意識すべき顕著な相違のいくつかを示している。

最もよく知られた ILSA は、国際教育到達度評価学会（IEA）および経済協力開発機構（OECD）のコア調査である。

- IEA 国際数学・理科教育動向調査（TIMSS）
- IEA 国際読書力調査（PIRLS）
- IEA 市民性教育国際調査（ICCS）
- IEA 国際コンピュータ及び情報リテラシー調査（ICILS）
- OECD 生徒の学習到達度調査（PISA）

Source: OECD 2019; IEA 2019

表1：IEA 調査と OECD の PISA 間のアプローチの違い

特徴	IEA 調査	OECD PISA
調査方針	学校で教えられることと学習に関連する要素を測定評価する。	義務教育終了までに学生が取得する一定のスキルセットを測定評価する。
内容の選択	参加国のカリキュラムを分析する。続いて参加国は、それぞれの国の関心事が確実に考慮されるようにしつつ、共同で評価のための枠組みとテストマテリアルを開発する。	OECD 選定の専門家たちは、学生が、将来使用するために習得すべきと考えるスキルを決定し、それに応じて調査手法をまとめる。
調査対象群の選択	サンプルはグレードベース（グレード4，8、12）で、カリキュラムの構造を反映し、教科担当教師と直接リンクさせる。	サンプルは年齢ベース（PISA については15歳）で、学校パスウェイまたはグレード分布とは無関係に、世代を評価する。

1. https://www.iea.nl/studies/additionalstudies/rosetta

さまざまな組織

調査そのもの以外に、IEA および OECD という2つの組織は多くの点で異なっている（表2参照）。最も根本的な違いは、調査結果と政策立案との関係における責務である。OECD の職務範囲は元来経済セクターに根差しており、教育の枠を超える。OECD の責務の1つは、実施すべき政策に関する提言をOECD加盟国に対して行うことである。一方、IEA は教育面の学術研究に根差した協会であり、メンバーに対する提言をまとめることはIEAの職務ではない。むしろ、個々の国が自身の状況に即した適切な政策を策定するために、よって立つことのできるエビデンスを提供することを責務としている。以下の表は、IEA と OECD の組織上の違いの一部を示している。

表2：IEA と OECD の組織上の違い

特徴	IEA	OECD
組織の種別	非政府協会（国内メンバーを擁する）。	政府組織。
ILSA の実施方法	自身の科学的ネットワークのパートナー団体と協調して調査を実施。	調査を立ち上げ、4年ごとの調査のための入札を行う。さまざまなタスク（例：枠組み開発、サンプル抽出、調査プラットフォーム）を区別する。
レビュープロセス	調査の発表物はすべて、厳格にピアレビューされる。	内部で作成およびレビューされ、参加国の役員メンバーによって精査される。
参加	どの国でも参加できる。調査参加の要件はない。	フラッグシップである PISA 調査は、当初OECD加盟国を対象としていたが、現在は、どの経済圏からでも参加できる。
意思決定	最終的な内容は、参加国の調査コーディネーターによって決定され、それぞれのコーディネーターが平等な発言権を持つ。	組織構造は、OECD メンバーシップの本来の中心性を反映し、投票は OECD メンバーおよび特定のパートナー経済域に制限されている。
費用	各国は、国際協力費用について均等に分担する。	加盟国が負担する拠出金は、自身の GDP に応じて異なる。庁舎に参加する非加盟国は定額を支払う。

共通の手法

組織とアプローチの面で異なるものの、ILSAを実施する際の手順や手法は極めて類似している。これまで60年以上にわたって開発および改善されてきたそれらの手順や手法は、多くの国における地域レベルの ILSA および国内の評価プログラムの手法の進展に寄与してきた。世界各地の専門知識を結集させ、すべての主要 ILSA において、調査 プロセスにおける各ステップ（サンプル抽出ルール、翻訳プロセス、実地調査手順、心理測定モデリングなど）の目的を達成するために慎重に規定された高い品質基準が義務付けられている。こうした手法は、ILSA に馴染みのない人にとっては極めて複雑に思えるかもしれないが、調査の透明性を確認できるように、データにはそれを支持する詳細な技術報告書が加えられ、どのILSA データも研究者はインターネットから自由にダウンロードできるようになっている。さらに IEA および OECD は共に、データの二次分析の推進を後押しする数多くのリソースを生み出している。こうした二次分析は、コンテキストデータが存在することで可能となり、豊かなものとなる。最近の事例としてICILSを挙げることができる。この調査では、コンテキストデータによって学生のデジタルリテラシーについて十分な知識を得ることができた。

評価に当たっての限界

ILSA 調査結果の理解と報告にあたって検討すべき重要な調査上の限界がいくつかある。まず、ILSA はすべて横断的研究である。すなわち、特定の人口群を対象に、ある時点での教育上の成果が測定評価されるのである。さらに、TIMSS、PIRLS、PISA など一部の ILSA は、経時的な動向を報告できるが、個々の学生を追跡する調査ではない。結果として、実行不能ではないものの、学生の学習到達度についての因果的結論を導くのは難しい。また、多くの政策立案者が、教育制度の改善のために何を変化させるかに関する具体的な提言を探ろうとしても、そのようなデータには明らかに限界がある。一部の研究者たちは、高度なモデルを用いて因果関係を確立しようとするが、そういったモデルは強引な仮定に基づいており、ILSA データを使ってモデルを確立することは、絶対不可能ではないにせよ困難であるとIEAは考える。

もう１つのILSA の限界とは、調査が、個々の学生の学習到達度および個々の学校の学習成果を測定することではなく、教育上の成果と教育制度内の背景情報との関係が反映されるように設計されていることである。従って評価は、学校、教師、学生にとってあまり重大な意味を持たない。しかしながら、評価を重要視しないことで、学生や学校にかかる調査のストレスが低くなるという利点がある。事実 ILSA は、重要視される評価と比べ、試験に要する時間かなり短く、また一国の人口群に対してごくわずかな人々 に対して実施されるに過ぎない。

最後に取上げる調査の限界は、ILSA が対象とする学習分野が、学校で教えられている内容を網羅していないという点である。例えば IEA の TIMSS は、グレード4と8の数学および理科の学習到達度を評価する調査である。重要な科目ではあるが、これらの科目の学習到達度だけでは、教育制度全体の健全性に関する信頼できる指標とはなり得ない。また、ILSAが国の教育政策に好ましくない影響を与えているという批判もされてきた。その批判は、教育に関する国の取り組みが、ILSAの評価対象となる科目の学習到達度が改善する方向で教育カリキュラムを組もうとする傾向になっているとの認識に基づいている。従って、教育制度の全体的な品質測定手段として 、ILSA の調査結果を考慮することには注意が必要である。むしろ、ILSAの調査結果は、特定の学科に関して学生が何を知っていて、何ができるか、そしてどのように国内の学生が、海外の学生たちと比較を行うかについての重要な指標とされるべきである。

ILSA が重要である理由

そのような限界はあるが、ILSA は教育制度の改善にあたって不可欠のツールである。ILSAは、データから読み取れる情報を使用し、堅固な国際的な枠組みの中でシステムを経時的に監視することができるという点で特異なものであり、どの国の統制下にもなく、どんな政治システムからもほぼ独立している。さらに、データは一般公開されているため、世界中の研究者たちはデータを検証し、自身の研究上の疑問や課題のために活用することができる。

多くの国で、ILSA の結果に由来する研究は、教育制度がどのように機能するかについての理解を深め、強固で信頼の置ける エビデンスに基づく意思決定を促進してきた。さまざまなインパクト調査によれば、ILSA の結果が政策立案を支持するために活用されており（例：Breakspear 2012; Schwippert & Lenkeit 2012; Wagemaker 2013）、TIMSS および PIRLS Encyclopedias の報告にもある通り、多様な教育上の改善が ILSA による エビデンスによって促進されてきた。このように、ILSA のデータは正しく理解および分析されることで、政策決定および教育制度改善に向けての研究を後押しする有益な機会を提供するのである。

今後の展開

常に最新のメソッドを採用するILSA（直近では、コンピュータベースの評価への移行を行った）には、手法関連の新たな機会と課題が待っている。こうした理由から、ILSA を実施する組織は、国際的な評価の分野における研究調査を積極的に促進している。例えば IEA は、外部の研究者を対象に学術ジャーナル、カンファレンス、テーマ別報告書のスポンサードのみならず、評価分野のいっそうの進展を促進するために自身の研究チームの採用も行っている。より一般的には、IEA は刷新された戦略において、有望なトピックである「プロセスデータ」（評価に合格する際に学生が残すデジタルトレース）などを含め、研究とイノベーションを重視している。

ILSA 開発におけるもう1つのフォーカスは、より規模が大きく複雑な次元である「21世紀型スキル」と呼ばれるものの探究である。IEA は最近、こうしたスキルのマッピングを目的とするカリキュラム調査（21CS MAP）を立ち上げた。この基礎研究は、具体的な評価プログラムの開発に先駆け、学校で教わる内容（本来の、実施されるカリキュラム）に基づいたものになる。

結論

世界がますます互いに関連し合う中、我々は、ILSAによって他から学び、比較を通して自らをよりいっそう理解できると考える。IEA は、データの解釈と分析の方法について研究者や政策立案者を訓練し、調査結果についての詳細な報告書の作成や委託、ならびに興味深い調査結果の簡潔で分かりやすい四半期ごとの要約書の発表を通して、政策や研究コミュニティの支援に熱心に取り組んでいる。そうした活動は、「教育制度内や教育制度全般にわたる指導と学習の質を向上させることを目的として、教育慣行、プロセスおよび政策をより良く理解する」という IEA のミッションをサポートしながら実践されている。IEA 調査と報告はすべて、テストプロセスおよび調査や報告の厳格なピアレビューの完全な透明性を含む、社会科学研究の最も高い学術基準に準拠している。

著者紹介

DR DIRK HASTEDT

Dr Dirk Hastedt は、IEA の常任理事である。彼は、IEA の運営、調査、サービスを監督し、IEA の全体の戦略ビジョンを推進する。さらに彼は、メンバー国、研究者、政策立案者および教育セクター内の他の重要なプレーヤーたちとの強固な関係の構築と維持を行う。また Dr Hastedt は、IEA-ETS Research Institute（IERI）ジャーナルである Large-scale Assessments in Education の共同編集長も務める。

DR THIERRY ROCHER

Dr Thierry Rocher は、2018年10月の第59回総会において IEA 議長に選出された。それまで Dr Rocher は、常任委員およびフランスの総会代表者を務めていた。彼は現在、フランスの国民教育省（Ministère de l'éducation nationale）の学生評価局（Office for Student Assessment：DEPP）の長に就いている。

参考資料

Anderson, L.W., Ryan, D.W., & Shapiro, B.J. (1989). *The IEA classroom environment study*. Oxford, UK: Pergamon Press.

Breakspear, S. (2012). *The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance*. OECD Education Working Papers No. 71. Paris, France: OECD Publishing

Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A., & Walker, D.A. (1962). *Educational achievements of thirteen-year-olds in twelve countries: Results of an international research project, 1959, 1961*. Hamburg, Germany: UNESCO Institute for Education. https://unesdoc.unesco.org/ark:/48223/pf0000131437.

Husén, T. (1967). *International study of achievement in mathematics: a comparison of twelve countries*. Vol. I–II. Stockholm, Sweden and New York, NY: Almqvist & Wiksell and John Wiley.

Husén, T. (1983). *An Incurable Academic: Memoirs of a Professor*. Oxford, UK: Pergamon Press.

IEA. (2018). *60 years of IEA 1958–2018*. Amsterdam, Netherlands: Author. https://www.iea.nl/index.php/publications/60-years-iea-1958-2018.

IEA. (2019). IEA studies [webpage]. https://www.iea.nl/studies/ieastudies

Lockheed, M. (2011). Reflections on IEA from the perspective of a World Bank official. In C. Papanastasiou, T. Plomp, & E. Papanastasiou (Eds.), *IEA 1958-2008: 50 years of experiences and memories, Vol 2*. Nicosia, Cyprus: Cultural Center of the Kykkos Monastery.

Lockheed, M., Prokic-Bruer, T. & Shadrova, A. (2015). *The Experience of Middle-Income Countries Participating in PISA 2000-2015*. Paris, France: PISA/OECD Publishing.

OECD. (2019). PISA: Programme for International Student Assessment. http://www.oecd.org/pisa/

Schwippert, K., & Lenkeit, J. (Eds.). (2012). *Progress in reading literacy in national and international context: The impact of PIRLS 2006 in 12 countries*. Münster, Germany: Waxmann Verlag GmbH

Travers, K., & Westbury, I. (1989). *The IEA Study of Mathematics I: Analysis of the Mathematics Curricula*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).

Wagemaker, H. (2013). International large-scale assessments: From research to policy. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, p. 11. Florence, KY: Chapman and Hall CRC.

その他のリソース

Hastedt, D. (2016, November 7). The history and development of international assessments [Podcast]. FreshEd Podcast 49. http://www.freshedpodcast.com/dirkhastedt/.

Lockheed, M., & Wagemaker, H. (2013). International Large-Scale Assessments: Thermometers, whips or useful policy tools? *Research in Comparative and International Education, 8*(3), 296–306.

Singer, J.D., Braun, H.I., & Chudowsky, N. (2018). *International education assessments: Cautions, conundrums, and common sense*. Washington, DC: National Academy of Education.

IEA COMPASS

IEA について
国際教育到達度評価学会は、IEA として知られており、各国の研究機関や政府機関から成る独立系の国際コンソーシアムで、本部はアムステルダムにある。その主たる活動目的は、教育制度内および教育制度全般における政策と実践の効果に対する深い知識を得ることを念頭に、教育的到達度に関する大規模な比較研究を行うことである。

Copyright © 2020 国際教育到達度評価学会（IEA）
禁無断転載。著作権者の書面での許可なく、電子、静電、磁気テープ、機械、複写、録画や記録など方法や手段によらず、本出版物の複製、検索システムへの保存、伝送は禁じられている。

ISSN: 2589-70396

本出版物のコピーは以下から取得可能である：

IEA Amsterdam
Keizersgracht 311
1016 EE Amsterdam
The Netherlands

Eメール：secretariat@iea.nl
ウェブサイト：www.iea.nl

Thierry Rocher
IEA 議長

Dirk Hastedt
IEA Executive Director

Andrea Netten
アムステルダム理事

Laura Cheeseman
コミュニケーション担当責任者

Compass 編集者
David Rutkowski
インディアナ大学
(Indiana University)

フォローする：
@iea_education
IEAResearchInEducation
IEA

本出版物を以下のように引用願います：
Rocher, T., & Hastedt, D. (2020, September). *International large-scale assessments in education: a brief guide* IEA Compass: Briefs in Education No. 10. Amsterdam, The Netherlands: IEA