中学校の先生・生徒対象

TOEFL Primary®の
スコアレポート活用法の検討
――テストについて知ることを通して――

大阪教育大学　英語教育部門　　准教授
篠崎　文哉

1. はじめに …… p.1
2. テストの種類と特徴 …… p.1
3. テスト設計 …… p.2
　3.1 信頼性 …… p.2
　3.2 妥当性 …… p.3
　3.3 実用性 …… p.4
　3.4 真正性 …… p.4
4. TOEFL Primary®のテスト結果の活用例 …… p.5
5. おわりに …… p.7

【著者紹介】

篠崎　文哉（Fumiya Shinozaki）　大阪教育大学 多文化教育系教員養成課程英語教育部門　**准教授**
ハワイ大学マノア校（Second Language Studies，学士）、大阪教育大学教育学研究科英語教育専攻（教育学，修士）。
大阪教育大学附属天王寺中学校教諭を経て、現職。専門領域は、英語教育学（プレゼンテーション、ICT活用）。
近著に、「端末数の差と共同編集機能の有無が英語グループプレゼンテーション準備への参加態度に及ぼす影響」（コンピュータ&エデュケーション）や「中学校での英語プレゼンテーションの質疑応答時に生徒が抱える困難点と活動環境の調整が与える影響―アンケート結果からの考察―」（英語教育研究）などがある。

【参考文献】

● 静哲人 (2015).『英語テスト作成の達人マニュアル』大修館書店.
● 正頭英和 (2018).『6つのアイデア×8の原則で英語力がぐーんと伸びる!英語テストづくり&指導アイデアBOOK』明治図書出版.
● 白畑知彦・冨田祐一・村野井仁・若林茂則 (2008).『英語教育用語辞典』大修館書店.
● 横川博一「TOEFL Primary®やTOEFL Junior®は『対策』が可能か―問題の本質を見抜いて確かな英語運用能力を―」
● Kumon Institute of Education Co., Ltd. (2018).「TOEFL Primary® Step2 サンプルテスト」
https://gc-t.jp/sample_test/primary/step2/lp4-2.html
● Kumon Institute of Education Co., Ltd. (2022). TOEFL family research program.
https://www.ets.org/toefl/research/reliability-validity.html
● MetaMetrics Inc. (2022). LEXILE framework for reading. https://lexile.com/educators/understanding-lexile-measures/

1. はじめに

中学生のみなさんは「テスト」と聞くと、何を思い浮かべるでしょうか。おそらく学校で受けているテストが最初に浮かんだのではないでしょうか。では次に、英語の授業で行われるテストにはどのようなものがあるか考えてみてください。カテゴリーなどを無視して列挙してみると、例えば、中間テスト、期末テスト、実力テスト、小テスト、単元テスト、単語テスト、文法テスト、音読テスト、発音テスト、リスニングテスト、スピーキングテスト、リーディングテスト、ライティングテストなどあり、もっと細かく区切れば無制限に出てくるかもしれません。

学校で配付される年間行事予定表には、中間テストや期末テスト、実力テストの日程が書かれていると思います。中間テストと期末テストはいわゆる定期テストです。では定期テストと実力テストを比較してみると、何か違いはあるでしょうか。最初に思い浮かんだのはきっとテスト範囲の有無でしょう。定期テストでは、出題範囲が事前に生徒に提示されている一方、実力テストと呼ばれるテストを受ける時、出題範囲は提示されないでしょう。

このように、学校で受けるテストには種類があることがわかります。テストを作成する教員だけでなく、テストを受ける生徒のみなさんも、テストについて理解を深めることで、次への英語学習につなげられるかもしれません。範囲を広げ、外部テストも含めて考えてみましょう。

2. テストの種類と特徴

定期テストと実力テストを簡単に比べてみましたが、ここではもう少し詳しくテストの種類とその特徴を見ていきたいと思います。

定期テストは、いわゆる到達度テスト（achievement test）に分類されます。到達度テストとは、「一定の学習期間で特定の学習目標にどれだけ到達したのかを測定するテスト」のことです（白畑・冨田・村野井・若林, 2008, p.7）。また、目標準拠テスト（criterion referenced test）でもあります。これは、「学習および指導の目標を設定し、それらの到達すべき目標基準に従って学習成果を測るテスト」のことです（白畑ほか, 2008, p.84）。一方、実力テストは、熟達度テスト（proficiency test）に分類されます。熟達度テストとは、「特定の学習目標やシラバス（syllabus）、学習期間、学習方法などから離れて、学習者が現時点で、どの程度の言語運用能力を持っているのかを測定するテスト」のことです（白畑ほか, 2008, p.245）。また、集団準拠テストでもあります。これは、「個人の成績を、そのテストを受験した受験者全員からなる集団の標準（norm）と比較することによって評価するテスト」のことです（白畑ほか, 2008, p.207）。

TOEFL Primary®は、この熟達度テストに含まれます。TOEFL®には、TOEFL Primary® STEP1、TOEFL Primary® STEP2、TOEFL Junior®、TOEFL ITP®、TOEFL iBT®があります。Primary®からiBT®にかけて、徐々に難易度が上がっていきますが、定期テストのようにある学習期間の中で、どれほど学習した内容が身についているかを確認することを目的としているわけではありません。例えば、TOEFL Primary® STEP2は、世

界約50の国と地域の英語を母語としない小中学生を主な受験対象としており、CEFR（外国語の運用能力をレベル別で示す国際標準規格。Pre-A1、A1、A2、B1、B2、C1、C2という段階があり、A1が基礎レベル。）でいうA1 ～ B1あたりに求められる熟達度を測定するテストとなっています。ですので、中学校で学んでいるレベルに近いと思われます。ただし、熟達度テストですので、定期テストのように、「この単語リストから出題される」「この文法事項を中心に学習しておく必要がある」といったような明確な範囲はありません。とはいえ、公式問題集はありますので、おおまかな雰囲気は事前につかめるかと思います。

3. テスト設計

定期テストにしろ、TOEFL Primary®にしろ、思いつきで問題を作成しているわけではありません。テストをするというのは、つまり何らかの反応を引き出して評価するということです。テスト作成に際して、まずはテストの全体像を考えることが普通です。例えば、どういう力を測定したいのか、どのように測定するのか、測定するのに何問必要か、テスト実施後にどのように採点するか、などを検討しています。ここでは、TOEFL Primary® STEP2を例にもう少し詳細にテストについて考えてみましょう。

3.1 信頼性

テスト検討をする観点の一つに、信頼性というものがあります。簡単に言うと、そのテストで何度測っても同じような結果が出るかということです。テスト受験者からすると、複数回同じテストを受けたとき、だいたい同じ結果が得られるかということです。

もしあなたが、同じときに2回TOEFL Primary® STEP2を受けたとして、片方のスコアが200点で、もう片方が230点だったとしたら、どう思いますか。たまたま運よく高得点が取れたと思うかもしれません。では、5回受けてすべての点数が極端にバラバラだったらどうでしょうか。おそらくテストの精度を疑うでしょう。実際には、人ですから、その時々によって頭の回転の調子が良いということもあれば、体調が優れないということもあります。連続でテストを受ければ、当然疲れます。わからない問題に対して、当てずっぽうで回答することもあるかと思います。様々な要因が絡みますので、全く同じスコアが連続して続くということはほとんどないでしょう。しかしながら、そのぶれ幅が極力小さくなるように、TOEFL®などのテストは丁寧に設計されているわけです。

信頼性を高める工夫の一つとして、項目の数を増やすということが挙げられます（静, 2015）。TOEFL Primary® STEP2は、リーディングとリスニングの2部構成となっており、それぞれ37問と39問あります。セクションごとで測る力が異なりますが、各セクションに複数の問題が含まれています。TOEFL Primary®を受験すると、スコアレポートが発行されます。もし受験したテストの信頼性が低かったら、そのスコアレポートも信用できません。逆に、信頼性が高いと、自分の強みと改善点を知ることができます。TOEFL®のテストを作成しているETSは、信頼性や妥当性を高めるための努力を重ねており、多くの研究者の調査も公表しています（ETS, 2022）。

3.2 妥当性

では、妥当性とは何でしょうか。TOEFL Primary® STEP2では、以下のような力を測ろうとしています（Kumon Institute of Education Co., Ltd, 2022）。

リーディング

- 複雑な文書やあまり馴染みのない単語を含む文章を理解することができる。
- メニュー、スケジュール、ポスターなどから情報を見つけ理解することができる。
- 一連の指示を理解することができる。
- 約250語からなる物語を理解することができる。
- アカデミックな話題について書かれた文章を理解することができる。
- 長い文章中から様々な情報を見つけ、関連付けることができる。
- 結果を推論し、導き出すことができる。

リスニング

- あまり馴染みのない単語を含む会話やメッセージを理解することができる。
- 約250語からなる物語や簡単なアカデミックな文章を理解できる。
- 先生の指示、メッセージ、物語、対話、会話および簡単でアカデミックな文章などの情報を関連付けることができる。
- 会話を聞いて結果を推論し、導き出すことができる。

ここに挙げられていることができるかどうかを測るために、問題が作成されているということです。テストが測ろうとしている知識や能力をしっかり測ることができているとき、そのテストは妥当性が高いと言います。

さきほど、信頼性が高いと強みと改善点を知ることができると述べましたが、実はそれは妥当性が担保されていることが前提となります。何回かテストを受けて、仮にどれも同じようなスコアが算出されたとしても、測るべき力が測られていないとそのスコアを適切に解釈できません。

ここでTOEFL Primary® STEP2のリーディングのサンプル問題（Kumon Institute of Education Co., Ltd, 2018）を一つ見てみましょう。

リーディングのサンプル問題

There are many ways to communicate. People communicate with other people not only by speaking and writing but also by making faces and moving their arms and heads. Animals communicate too. For example, blackbirds make loud noises to communicate danger. Bees use a special dance to explain where to find food. Elephants move their ears up and down when they are excited.

Question
What is the reading about?
(A) How animals and people are different
(B) How animals and people communicate
(C) How animals and people move their bodies

When do bees dance?
(A) When they are excited
(B) When they are in danger
(C) When they are telling others about food

コミュニケーションというテーマであり、その中身もアカデミックなものとなっています。また、設問も大意把握をするものや、特定の箇所を読み取るようなものがあり、文章が理解できているかを測る問題となっていることがわかります。つまり、先ほど示した測りたい力と合致しているということです。

3.3 実用性

ただ、どれだけ妥当性の高い問題を作ったとしても、そのテストが実施可能なものでなければ意味がありません。テストの作成や実施、採点などが現実的に可能なのかについて検討する観点が実用性です。例えば、信頼性を高める工夫の一つとして、項目の数、つまり問題数を増やすということを挙げましたが、では1回のテストに何千もの問題が含まれているとしたらどうでしょうか。妥当性の高い問題をそこまで多く用意するのは困難ですし、受験者からするとすべて回答するのにどれだけの時間と労力がかかるのだろうと心配することでしょう。

現実には、テスト実施にかけられる時間の制約があることが多いですので、その中で信頼性と妥当性をできる限り高める努力をしています。つまり、テスト作成者、採点者、受験者など様々な観点を考慮してテストが実施されていることになります。受験者のみなさんには、スコアレポートをもとに学習の成果を振り返ってほしいですが、絞られた問題数であっても強みや改善点が見えてくることかと思います。

3.4 真正性

英語の先生は、英語学習者に英語を使えるようになってほしいと思っています。使えるようになるためには、語彙力や文法力といった基礎力が必要です。語彙問題や文法問題を解くことで、身についている力を確認することがあります。

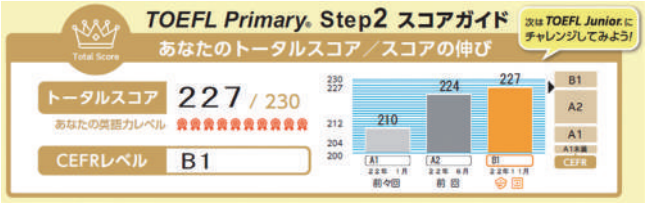
しかし、問題設定によっては、あまり現実世界では出くわさないものがあります。例えば、正頭（2018）は、「『駅で流れるアナウンスを聞き取り、何番ホームの電車に乗ればいいかを答えなさい』というようなリスニング問題は真正性が高く、『空欄補充』のような現実生活の中ではおそらくないであろう状況は真正性が低いと表現できます」（p.19）と説明しています。つまり、真正性とは現実に近い状況の問題設定ができているかを指しています。

横川博一先生のレポートでは、TOEFL Primary® STEP1の問題集にある問題を例に、ハサミや図形など身の回りにあるもの、つまり「生活語彙」が多く出現していることに触れられています。さらに、リスニング問題を取り上げ、そこでは小中学生にとってはやや高度な文法事項が含まれているが、自然なやり取りになっていることが説明されています。言葉を引用すると、「一見やさしそうに見える問題も、語彙や文法の単なる暗記や付け焼刃の対策では到底太刀打ちできるものではないようです。問題を少し分析しただけでも、子供たちが学校生活や社会生活で経験するごく普通の場面で、自然なやりとりの中で外国語としての英語を生きたことばとして学ぶことが望ましいのだと訴えかけているように思います。」（p.5）と述べられています。TOEFL Primary®では、まさしく真正性が担保された問題が多く含まれているということです。

真正性が低くても学習には効果的なものは多くありますので、学習の過程として必要だと思います。ただ、そこに留まらず、英語を実際の場面で「使う」ことを念頭に学習を進めてほしいと思います。

4. TOEFL Primary® のテスト結果の活用例

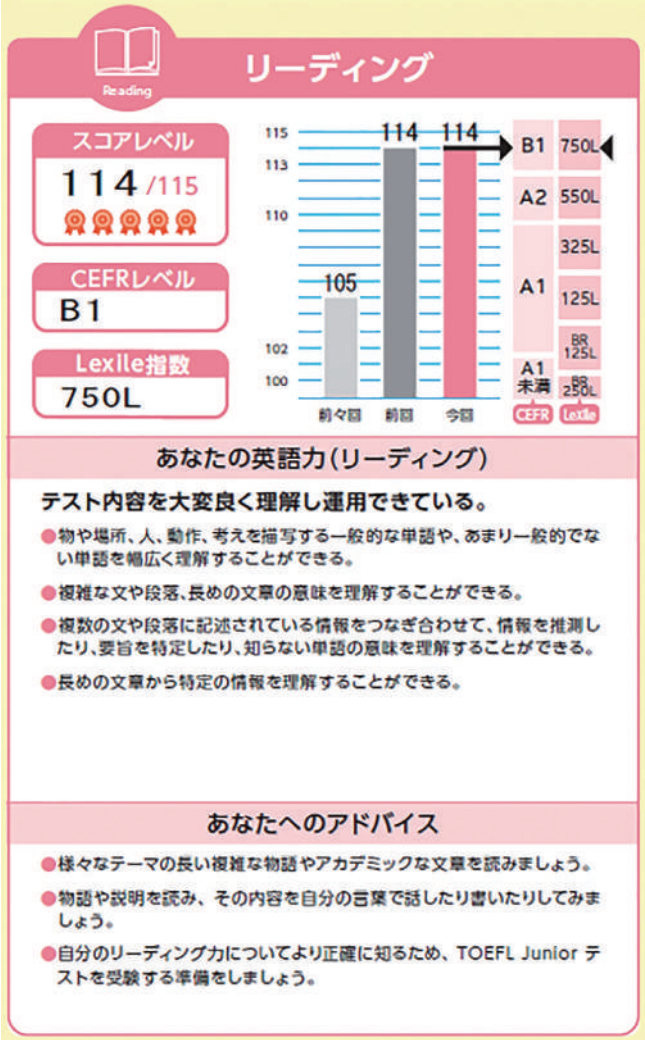
様々な観点からテストについて理解を深めてきましたが、テストの質が高ければ、その分、テスト結果もしっかり見る意義が出てきます。ここからは、TOEFL Primary® STEP2のスコアレポート例を参照しながら、その見方を確認したり、活用法を検討したりしていきます。



スコアレポートの最初には、トータルスコアとスコアの伸びが書かれています。例では、230点中227点だったことがわかります。非常に高いスコアです。定期テストでもそうですが、テストを受ける限りは、高い点数を取りたいと思いますし、実際に高い点数が取れると嬉しいです。しかし、テストを受けて、結果が返却されて点数を見たらそこで終わりというのは、テストの使い方としてはもったいないのです。特にTOEFL®では詳細にテスト結果を示してくれますので、細部まで確認した上で、次の学習につなげられるよう自分なりに分析するべきです。

では、トータルスコア以外のところに目を向けてみましょう。今回は、CEFRでいうところのB1というレベルでした。これまでに複数回受験していれば、過去のスコアとの比較がなされます。上記の例では、2022年1月には210点（A1）、5か月後の6月には224点（A2）、そして今回11月では227点（B1）となっています。グラフで数値の伸びを感じると、モチベーションにつながると思います。

どのようにスコアが伸びたか見てみましょう。1回目から2回目で大きくスコアが伸びていることがわかります。継続的な努力の結果だと考えられます。一方、2回目から3回目（今回）では、あまり伸びが見られません。努力が足りなかったのでしょうか。それは受験者本人が最もよく知るところではありますが、努力不足とは限りません。満点またはそれに近いスコアを取得した場合、あなたの力を測りきれていない可能性があります。小学1年生が最初に習う足し算や引き算の問題を中学生が解いたら100点を取れると思います。100点ですので、もう学ぶことはないのでしょうか。当然ですが、そうではありません。中学レベルの数学の問題であれば、100点を取るのはなかなか難しいと思います。仮に70点だったとしたら、できている部分は多いですが、確実にできていると言えない部分もそれなりにありそうです。このことは、スコアの算出方法が異なるTOEFL Primary®でも同じことが言えます。つまり、なるべく自分のレベルに合ったテストを受けた



方が、自分のことをより正しく知ることができ、次の学習に生かしやすくなるのです。そういうこともあり、実際に例の中でも「次はTOEFL Junior®にチャレンジしてみよう!」と次の段階に移ることを勧められています。なお、こういったスコアの伸びは、前述した通り、信頼性や妥当性が高くなければ比較する意味はあまりありません。

次に、リーディングの結果を見てみましょう。さきほどと同じように、CEFRレベルや過去のスコアからの変化が掲載されています。これらに加えて、Lexile指数というものがあります。MetaMetrics Inc.(2022)では、Lexile指数(Lexile Framework for Reading)は、読解力とテキストの複雑さの両方を同じ発達段階尺度で測定する科学的アプローチだと説明されています。自分のLexile指数を知ることができれば、そのレベルに適した本などのテキストを探して読み、リーディング力を効率よく伸ばすことができます。自分のLexile指数のマイナス100Lからプラス50Lまでが目安となっています。今回の例では、自分のLexile指数は750Lですから、選ぶテキストは650L～800Lとなります。テキストを探す方法は色々ありますが、スコアレポートでは、以下のような説明があります。

5 Lexile®指数(レクサイル指数 リーディングのみ)

読書能力と英語の本の難易度を表わす指数です。
下の専用サイトからあなたの「読む力」にぴったりの英語の本をみつけることができます。
※Lexile®はアメリカMetaMetrics社の商標です。

Lexile *TOEFL Primary®*/*TOEFL Junior®* 専用サイト http://toefljunior.lexile.com/

BR●●Lとは?…数字の前にBRがついたものは初級読書者(Beginning Reader)をあらわします。

この専用サイトでは、自分のLexile指数を入力し、興味のある分野を選択すると、それらに応じた記事や本を検索することができます。そこで、記事を読んだり、本を購入したりすることができます(ただし、本当に自分のレベルに適しているかは、テキストの中身を実際に少し見て確認してください)。テストスコアに加え、Lexile指数の変化にも着目し、リーディング力向上に役立ててほしいと思います。

さらに、「あなたの英語力(リーディング)」では、今回のテスト結果から、現在リーディングにおいて何ができるのかを示してくれています。自分の能力を客観的な記述から省察し、分析することは学習の過程において非常に大切です。何ができて、どこを改善する必要があるのかを建設的に検討することができます。

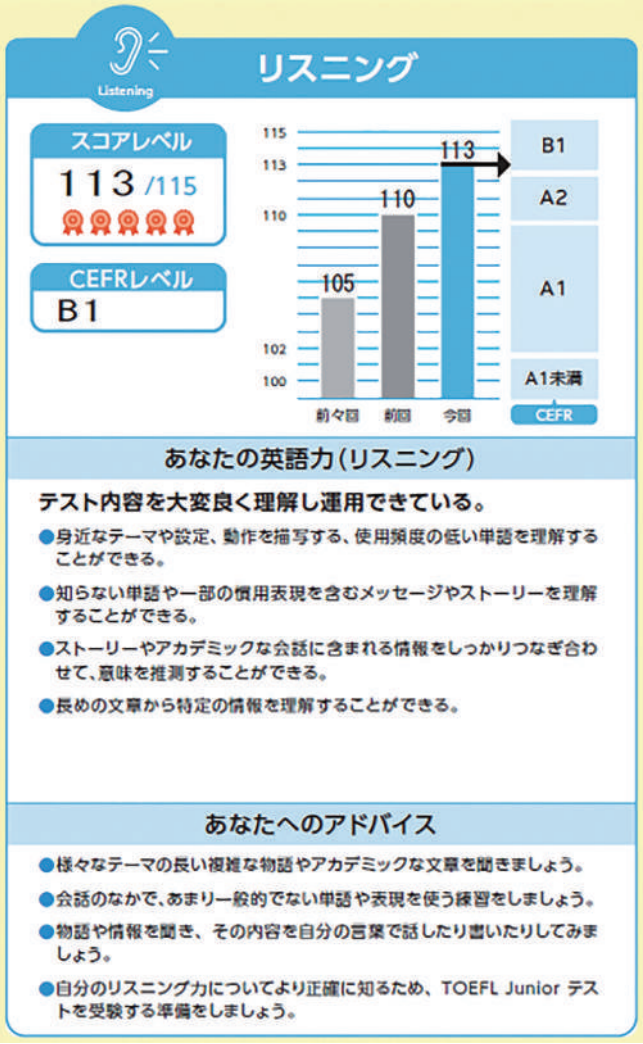
そのヒントとなるのが、「あなたへのアドバイス」の欄です。ここでは例として、「様々なテーマの長い複雑な物語やアカデミックな文章を読みましょう」ということについて解釈してみます。TOEFL®では、大きく「日常生活、コミュニケーション」「指示文」「アカデミック」が問題の構成要素となっており、そのうちのアカデミックの比率が、TOEFL Primary®よりTOEFL Junior®などの上位テストになるほど高くなっていきます。TOEFL®では、アカデミック・リーディングは「論文・教科書等専門書の記述スタイルやフォーマット、語句を理解する。」とされています。そもそもTOEFL Primary®は、「主に、小学生、中学生の日常生活での英語コミュニケーションや、学校の授業や教科書など、親しみのあるシチュエーションでの英語能力測定」をねらいとしていますので、受験者のボリュームゾーンにとっては、学校で学んでいることを適切に積み重ねて、さらに身の回りのことに関心を持ち、語彙力等を強化していくことで、高いスコアが取れると考えられます。しかし、今回の例では、114/115という極めて高いスコアでしたので、「自分のリーディング力についてより正確に知るため、TOEFL Junior®テストを受験する準備をしましょう。」という助言がなされているように、より高いレベルの学習をしてはどうかと提案されています。

様々な学習方法がありますが、日常的に文章を読むというのは、やはり効果的です。学習全般に言えますが、継続させることが多くの人にとっての課題です。リーディング力も一朝一夕では伸びません。楽しいことや、力がついていることが実感できることが大切だと思います。学校の授業での学習に加え、授業外でも例えば多読を

行うのがよいでしょう。

続いてリスニングの結果も見てみましょう。こちらは、トータルスコアと見方は同じです。「あなたへのアドバイス」を見てみると、「会話のなかで、あまり一般的でない単語や表現を使う練習をしましょう。」というものが入っています。リスニング力の強化には、音声を聞く、聞き取ろうとする練習はもちろん効果的だと思いますが、それ以外にも実際に英語を使うというのも意味があると考えられます。特に、アドバイスにあるように、さらに高いレベルの単語や表現を自分が直接関わる文脈の中で使っていくことで、それらの使い方の定着につながると思います。それはハードルが高いと感じる人については、もしこれまで十分と言えるほど音読を行ってこなかったのであれば、ぜひ今からでも音読を継続してやってみてほしいと思います。音読は、数多くの研究によって言語学習の様々な方面に効果があると報告されており、リスニング力もその一つです。

ここではほんの一例しか挙げていませんが、いずれにしてもスコアレポートを丁寧に読み、次の学習へいかにつなげられるのかを自分なりに検討することが大切です。どのような学習方法が効果的なのかが自分ではわからない場合は、先生に尋ねるとよいでしょう。その際、自己分析した結果とともに伝えると、より具体的なアドバイスをもらえると思います。



5. おわりに

本レポートでは、テストについての理解を深めることで、スコアレポートの意味や活用法を考えました。定期テストでも同じことが言えますが、何点取れたかというのは受験者として重要である一方、そこだけを見て、結果を次の学習に活用しないのは非常にもったいないです。

定期テストとTOEFL Primary®は種類の異なるテストですが、TOEFL Primary®が学校で受けている英語学習と全く関係がないかと言えば、そうではないことは受験したことのある人なら実感できたかと思います。定期テストを利用して学習の積み重ねと振り返りを行い、その時点での実力をTOEFL®を用いて確認し、自分の分析を行ってほしいと思います。スコアレポートは、自分の分析に役立つだけでなく、具体的な学習につなげるための先生とのコミュニケーションツールにもなるでしょう。