Listening Comprehension

Progressive Achievement Test

TEACHER MANUAL

JULIET TWIST, HILARY FERRAL,
VERENA WATSON, JIM MCNAUGHTON,
SALLY ROBERTSON AND MAGDALENE LIN

NZCER PRESS
Wellington 2010

NZCER PRESS
New Zealand Council for Educational Research
PO Box 3237
Wellington 6140
New Zealand

ISBN 978-1-877398-70-4

Distributed by
NZCER Distribution Services
PO Box 3237
Wellington 6140
New Zealand
www.nzcer.org.nz

Contents

Tables

Figures

Acknowledgements

The authors and publisher of PAT:Listening Comprehension wish to acknowledge the foundation laid by the previous versions of PAT:Listening Comprehension. The authors of these—Neil A. Reid, Ian C. Johnston and Warwick B. Elley—are acknowledged for the enduring quality of their work, which has been used in New Zealand schools for over 30 years.

This revision of PAT:Listening Comprehension also reflects the influence of our Australian Council for Educational Research colleague Andrew Stephanou, who has provided advice and assistance throughout the series of recent projects to review and revise the New Zealand versions of the Progressive Achievement Tests. We also acknowledge the influence of our NZCER colleague Charles Darr, whose advice and assistance at all stages of the project have been invaluable.

The redevelopment of PAT:Listening Comprehension was very much a team effort. Other NZCER colleagues involved in the development process included: Paul Shih, Paul Lam, Rachael Kearns, Joanna Morton, Bev Webber and Sue McDowall.

We wish to thank the following people, who were involved in reviewing test specifications, texts and questions: Murray Gadd, independent literacy consultant; Lyn Bareta, Learning Media Limited; Martyn Reynolds, Wellington College; Mark Edgecombe, Hutt Valley High School; Ruth Jeffery, Wellington High School; Alistair Du Chatenier, Thorndon School; and Denise Hitchcock and Jessica Craig, Accent School Support, Victoria University of Wellington.

Thanks are due to the copyright holders, who have generously given permission for their work to be used in the tests. Full acknowledgement of them is given in the teacher scripts.

Finally, we acknowledge the hundreds of classroom teachers and thousands of students around New Zealand who participated in various phases of the redevelopment and standardisation processes.

PART A

The Progressive Achievement Test of Listening Comprehension

1 Introduction

The Progressive Achievement Test of Listening Comprehension (PAT:Listening Comprehension) form part of a series of standardised tests developed specifically for use in New Zealand schools. PAT:Listening Comprehension is primarily designed to help classroom teachers to determine the level of achievement attained by their students in listening comprehension. Properly interpreted, and supplemented by information on the past achievement and present circumstances of each student, the results from the tests will help teachers to make informed decisions about the kinds of teaching materials, methods and programmes that are most suitable for their students.

The revision of PAT:Listening Comprehension is based on a conceptualisation of listening as being fundamentally different from hearing[1] (Rost, 2002). Listening goes beyond the passive physiological process of hearing to involve paying attention. Once a listener is attending, there is the potential for them to engage in meaning making. However, a listener's ability to engage must always be selective due to the limitations of attention and short-term memory.

For the purposes of this test, listening involves two main types of active processing, in which listeners:

- categorise language in ways that enable them to make sense of it, such as recognising words, appreciating the relationships between words (grammar), and understanding the prosodic features of a text (the patterns of stress and intonation in a language).[2]
- relate language to ideas in their memory and objects in the real world.

Both types of processing sustain each other and are conducted simultaneously.

1.1 Features of PAT:Listening Comprehension

PAT:Listening Comprehension is designed for the assessment of students in Years 3 to 10. There are eight tests, each aimed at a specific year level. Each test can also be used at adjacent year levels to suit particular groups of students in different schools. The content of the tests has been designed and carefully reviewed by subject experts.

Scores on each of the tests can be converted to locations on a measurement scale: the PAT:Listening Comprehension (PATL) scale. The scale has been qualitatively described and allows progress to be measured from Year 3 to Year 10.

The scale scores are supplemented with normative information in the form of stanines. Student results for listening comprehension can be compared with the achievement of a national reference group at specific year levels.

The tests can be administered at any time during the school year. However, bear in mind that the data used in the normative reporting were collected at the beginning of the 2010 school year.

---

1 However we certainly recognise that the physiological process of hearing is necessary for effective listening. Note that if a student who has previously done well at *reading* comprehension assessments does not do well in a *listening* comprehension assessment, the possibility of hearing impairment is worth considering.

2 For more information on prosody, see pages 11, 64, and 71.

The items in each test are in multiple-choice format. Students' responses are recorded in the booklets for the Year 3 and 4 tests; separate answer sheets are used for recording students' responses for the Year 5 to 10 tests.

The tests are designed to take 40 minutes of testing time. Tests require about 10 minutes of additional time for administration.

The tests can be marked and scored by hand. This manual includes blackline masters of student reports that can be used to map a student's performance on each test. Alternatively, NZCER offers automated marking and reporting options (see www.nzcermarking.org.nz for more details).

2 About this edition

This edition of the Progressive Achievement Test of Listening Comprehension has been prepared by test development and curriculum staff at the New Zealand Council for Educational Research (NZCER). The new edition includes the redevelopment of PAT:Listening Comprehension, which was last revised in 1994 (by N. A. Reid, I. C. Johnson and W. B. Elley). Teachers and specialists in the teaching and assessment of reading from around New Zealand have reviewed the test specifications, test items and test forms.

2.1 Purpose of the test

PAT:Listening Comprehension assesses students' listening comprehension; that is, it assesses students' comprehension of texts read *to* them. Because the texts are read *to* as opposed to *by* students, teachers can obtain data on student meaning making that is independent of their ability to decode printed words. This focus reflects *The New Zealand Curriculum* (Ministry of Education, 2007) requirement that students make meaning of ideas or information in the listening mode; it also reflects our understanding of the intention of key competencies.

PAT:Listening Comprehension can be used to measure progress in listening comprehension over time. It will also provide teachers with data that will assist them to group students in ways that will best develop meaning making.

PAT:Listening Comprehension sits alongside PAT:Reading Comprehension (Darr, McDowall, Ferral, Twist, & Watson, 2008). The two tests assess comprehension in different ways, and together will give teachers a broader and deeper understanding of their students' thinking about text.

2.2 Changes you may notice

In writing this edition of PAT:Listening Comprehension we have kept some of the features of the previous edition while incorporating some changes, as follows.

- The test content has been updated: all the comprehension texts and their accompanying questions are new.
- The presentation of teacher scripts, test booklets and answer forms has been updated.
- All texts, questions and possible answers have been recorded on CDs and are read by professional actors.
- All texts are of the type typically used in classroom reading instruction (that is, there are no scripted conversations in this edition).

2.3 Test content

PAT:Reading Comprehension and PAT:Listening Comprehension both reflect the theoretical underpinnings of the reading standards presented in *The New Zealand Curriculum Reading and Writing Standards for years 1–8* (Ministry of Education, 2009). While PAT:Reading Comprehension relates to all three aspects of the act of reading as described in the standards: learning the code of written language; making meaning;

and thinking critically, with PAT:Listening Comprehension, attention is focused on the making meaning and thinking critically aspects.

PAT:Listening Comprehension assesses students' ability to construct meaning from an oral text. The content reflects current developments in the teaching and learning of listening comprehension. Each test contains a mix of narratives, poems and informational texts. Within each test there is a range of text types, text content, text length, text difficulty, question type and question difficulty.

Question types

Most questions use the following question format:

What does Yung like most about the Māori man in the shop?

A He is fighting the government.

B He is self-sufficient.

C He is confident.

D He is respectful. [Best response]

(Test 8, 'Māoriland', question 6, track 37)

However, a significant number use a format whereby a sentence "runs on" into the best response. For example:

Colin is best described as

A talented. [Best response]

B shy.

C wealthy.

D brave.

(Test 4, 'Colin Murdoch', question 5, track 26)

The questions in PAT:Listening Comprehension have been classified to give an indication of the kinds of cognitive processes or comprehension skills students are likely to call on in order to answer the question. We have categorised questions quite broadly because we believe this kind of approach will be most useful to teachers. Other approaches, such as that taken by Kispall (2008), offer a much more precise, but complex, categorisation.

Questions are categorised as either retrieval, local inference or global inference.

Retrieval questions (R)

These questions require the listener to retrieve clearly stated information from the text. In other words, listeners are required to use minimal inference.[3] Following are two examples of retrieval questions.

3 We have decided to categorise some questions as retrieval because we believe that doing so may provide teachers with useful data. However, it is important to understand that this has not been an easy decision and that categorising questions is not a straightforward business.

Indeed, some theorists would argue against the idea of retrieval questions. They would argue that categorising a question as retrieval implies a mistaken belief that meaning is sent directly from the mind of one person into the mind of another: the first person (for example, the narrator of a PAT:Listening Comprehension text) "sends" the message and the second (for example, the listener of a PAT:Listening Comprehension text) "retrieves" it exactly as the first intended. For these theorists, the listener *in all instances* would use their background knowledge, their "stock of mental stuff" (Reddy, 1993, p. 107), to *infer* the reader's meaning. According to this theory, there would be no retrieval questions.

Another argument against categorising questions as retrieval is that in order to find the best response, the listener must *infer* the relationship between the question, the text and the best response. Accordingly, all questions involve at least some inference.

Text: "Geckos are the only lizards that have a voice."

Question: Geckos are the only lizards that

A have a voice. [Best response]

B have velvety skin.

C live on mountain slopes.

D lay eggs.

(Test 1, 'Facts about Geckos', question 4, track 40)

To correctly answer this question, the listener matches "Geckos are the only lizards that" with the explicitly stated text "Geckos are the only lizards that have a voice."

Text: "One of their vehicles was on fire, illuminating the only pathway the men had to safety."

Question: What illuminates the only pathway the men have to safety?

A Machine gun fire.

B A vehicle's headlights.

C A burning vehicle. [Best response]

D The soldiers' torches.

(Test 5, 'Willie Apiata', question 3, track 19)

To correctly answer this question, the listener matches words and phrases used in the question, "What illuminates the only pathway" with similar words and phrases used in the text, "One of their vehicles was on fire, illuminating the only pathway the men had to safety."

Local inference questions (LI)

These questions require the listener to infer by drawing on pieces of information at the sentence and paragraph level. Following is an example of a local inference question.

Text:

"Halt!"

Ten men block our way, roadside vagabonds whose swords glimmer, dull and greedy.

Sensei says nothing, eyes measuring the man in front of him.

"Give us everything you have," the leader says, "or I will slice you where you stand. Those kids too."

Question: What are the ten men after?

A Valuables. [Best response]

B A fight.

C Hostages.

D Information

(Test 5, 'The Back Road', question 2, track 4)

To correctly answer the question, the listener infers information from one part of the text. They must infer from "Give us everything you have" that "everything" refers to valuables. They do this by inferring that vagabonds are associated in some way with dishonesty, and that yelling "Halt!" and blocking the way are typical actions of those who steal the valuables carried by travellers.

Global inference questions (GI)

These questions require the listener to infer overarching ideas by drawing on pieces of information situated across the text as a whole. Following are two examples of global inference questions.

Text: [The whole text.]

Question: What is the main purpose of this text?

A To make a prediction.

B To describe a personal experience.

C To summarise different viewpoints.

D To state an opinion. [Best response]

(Test 8, 'True Colours', question 5, track 26)

To correctly answer the question, the listener infers information from across the text as a whole.

Text: "Seagull, seagull / Riding high… / The men and the women / And girls and boys / Look far, far smaller / Than painted toys… / I like to glide / On my wings and stare."

Question: The seagull tells us what he sees from

A a park.

B a boat's mast.

C high in the sky. [Best response]

D the beach.

(Test 1, 'The Seagull', question 1, track 24)

To correctly answer the question, the listener infers information from across the text as a whole.

We have also categorised questions in two other ways: prosodic and multiple perspectives. Questions of these types also fall into the main categories of retrieval, local inference and global inference.

In Table 1 some questions are described as falling into two categories. For example, question 4 in Test 7 is categorised as "LI P", indicating it is a question that requires listeners to use local inference as well as prosodic features to answer it correctly. In the interests of clarity, the student reports (see p. 22) refer only to the main categories of retrieval, local inference and global inference.

Prosodic (P)[4]

These questions assess the student's ability to appreciate prosodic features—the patterns of stress and intonation in a language (stress is the relative emphasis given to certain syllables in a word; intonation is the variation in voice pitch). These questions require students to pay attention to the sound of the narrator's voice in order to comprehend emotions such as enthusiasm, frustration, delight, certainty and excitement. One of the reasons the texts in this edition are read by actors is because of their skills in regulating emotional effects in their speech.

Note that in written language we use punctuation to capture some of the stress and intonation of spoken language. (For more information on prosody, see Rost, 2002.)

Following are two examples of prosodic questions.

4 Note that only questions that require listeners to use prosody to a large extent are categorised as prosodic.

Text: "I told them all the ins and outs, I told them about my human subjects, to whom I even managed to give names and lives that sounded quite realistic."

Question: How does the narrator sound when he says, "I told them all the ins and outs, I told them about my human subjects, to whom I even managed to give names and lives that sounded quite realistic"?

A Guilty.

B Nervous.

C Excited. [Best response]

D Knowledgeable.

(Test 7, 'Caught Out', question 2, track 4)

To correctly answer the question, the listeners need to use the narrator's generally high but also fluctuating pitch and his fast and rather "breathy" delivery to infer his emotional tone as one of excitement.

Text: [The text as a whole.]

Question: How does the narrator sound?

A Confident. [Best response]

B Anxious.

C Tired.

D Cross.

(Test 3, 'Big Birds', question 1, track 28)

To correctly answer the question, the listeners need to use the narrator's steady speed of delivery, his slightly lowered pitch, and the lack of tension in his articulation to infer his emotional tone as one of confidence.

Multiple perspectives (MP)

These questions assess the student's ability to appreciate that text can be interpreted in more than one way. Although it can be difficult to use a multiple choice question to assess student thinking of this type, we have found that it is possible to do so when an ambiguous piece of text can be interpreted in two credible ways. There are only four multiple perspectives questions: one in Test 3, two in Test 5 and one in Test 7.

Following are two examples of multiple perspectives questions.

Text: The text itself gives no clues, although "While their actions may have provided some immediate solace …" may be used by listeners to understand the first part of the question; that is, that knocking the first soldier down might be described as "satisfying".

Question: Knocking down the first soldier they saw on the top deck might be described as satisfying. It might also be described as

A clever.

B dishonest.

C cautious.

D unfair. [Best response]

(Test 7, 'Fighting Prejudice', question 5, track 42)

To correctly answer the question the listener uses information from their background knowledge in order to decide on a credible alternative to "satisfying".

Text: "She was fun, my dizzy flake of a mother. I loved her passionately and I didn't care that I was the one who had to organise the running of our flat, who had to write out the cheques for the bills, make sure she didn't spend all the money before the next payday, get the washing done, drag her off to buy groceries. 'They're so boring, Kirby!' she'd cry. 'So's being hungry,' I said every week."

Question: Mum is described as "so alive". She might also be described as

A sensible.

B arrogant.

C irresponsible. [Best response]

D insensitive.

(Test 5, 'About My Mum', question 3, track 25)

To correctly answer the question listeners need to use information such as "my dizzy flake of a mother", information about the narrator taking responsibility for the running of the household, and that Mum describes the running of the household as "so boring, Kirby!" in order to decide on a credible alternative to "so alive".

Text types

PAT:Listening Comprehension uses short stories, extracts from novels, poems and various informational texts.

Content structure

Table 1 below outlines how the content for PAT:Listening Comprehension has been organised. The abbreviations in the right-hand column refer to the question types explained above; that is: R = retrieval questions; LI = local inference questions; GI = global inference questions; P = prosodic questions; and MP = multiple perspectives questions.

Table 1 PAT:Listening Comprehension—content structure

Test number	Text title	Text type	Number of questions and sequence of question types
Test 1	The Dawn Parade	Short story	LI P; LI; LI P; LI; LI; LI P; LI
	Dragonflies	Information report	LI; LI; LI; R; R
	A Boy for a Pet	Short story	R; LI; R; LI; LI; GI
	The Seagull	Poem	GI; LI; LI; GI
	Sione's First Day at School	Short story	LI P; LI; LI; LI; LI; LI; GI
	Facts About Geckos	Information report	GI P; LI; LI; R; LI
Test 2	The Best-Loved Bear	Short story	LI; LI; LI; LI P; GI
	The Microwave Oven	Factual recount	LI; LI; R; GI; GI
	The Gorse Fire	Poem	LI; LI; LI; LI; LI
	Dinosaurs in New Zealand	Information report	LI; LI; R: LI
	Car Trouble	Extract from a novel	LI P; LI; LI; LI; LI P
	Plants that Eat Insects	Information report	GI P; LI; LI; R; LI
	In and Out the Windows	Poem	LI; LI; LI
	It's Raining Fish	Information report	R; R
Test 3	To the Rescue	Extract from a novel	LI; LI P; LI; LI P; LI; LI P; GI
	Joy Cowley	Information report	LI; R; LI; LI; GI
	Beans	Short story	R; LI; R; LI; LI; LI P; GI
	Night Countdown	Poem	LI; GI MP
	Big Birds	Information report	GI P; R; LI; LI; R; R
	Animal School	Short story	LI; LI P; R; LI; LI;
	Cheese Mould	Information report/instructions	LI; LI; GI

Table 1 PAT:Listening Comprehension—content structure (continued)

Test number	Text title	Text type	Number of questions and sequence of question types
Test 4	Sione Wakes Up	Short story	LI; LI; LI; LI; LI; GI; GI
	Tsunami	Informational report/factual recount	LI; LI; LI; LI; LI
	The Windy Night	Poem	LI; GI; GI
	Pass It On	Short story	GI; GI; LI; GI; GI
	Colin Murdoch	Information report	LI; LI; LI; LI; GI
	Roger is Boring	Short story	GI P; GI; LI; LI
	Vilavilarevo	Information report	GI; R; R; R; LI; R
Test 5	The Back Road	Extract from a novel	R; LI; LI P; LI; LI P; GI
	Pākoti Rouru—A Cook Islands Ceremony	Information report/factual recount	R; LI; R; LI; LI; R
	Willie Apiata	Extract from a novel	LI; LI; R; R; LI
	About My Mum	Poem	LI; LI P; GI MP; LI P; GI
	Sun o (2)	Instructions	GI P; LI; GI; LI
	Handling Chicken	Factual/personal recount	GI P; LI; R; GI
	Tā Moko		LI; LI MP; LI; R; GI
Test 6	Tunutunu Ram and The Dude	Short story	R; LI; LI; LI; GI
	Hillary Tribute	Factual recount	R; LI; LI; LI; LI
	Apples	Extract from a novel	LI; LI; LI; LI P; GI; GI; GI
	Don't Be Wet	Poem	LI; GI
	Early Fabrics: Nature's Materials	Information report	LI; LI P; R; LI
	Tuatara: Our Living Ancient Taonga	Information report	LI; R; LI P; LI; LI; GI
	Tangiwai Rail Disaster	Factual recount	LI; R; LI; LI; R; LI
Test 7	Caught Out	Extract from a novel	LI; LI P; LI; LI P; LI; GI
	Japanese Whaling	Argument	LI P; LI; LI; LI; GI; GI
	Freya	Extract from a novel	LI; LI; LI; LI; GI
	Milking Before Dawn	Poem	LI; LI; LI
	A Talking Ape	Information report	GI P; R; LI; R; LI
	Farmhand	Poem	LI; LI; LI; LI
	Fighting Prejudice	Factual recount	R; LI; LI; LI; LI MP
Test 8	Journey	Extract from a short story	LI; LI; LI; LI
	New Zealand Invader	Information report	R; R; LI; LI; LI; LI
	The Voice	Extract from a novel	GI; LI; GI; LI; LI; LI
	True Colours	Argument	LI P; LI; LI; LI; GI
	The Saxophone Teacher	Extract from a novel	LI; R; LI; LI; LI P; LI
	High Country Weather	Poem	LI; GI
	Māoriland	Extract from a novel	LI; LI; LI; LI; LI P; LI

2.4 Choosing a PAT:Listening Comprehension test

Each of the PAT:Listening Comprehension tests has been designed with a specific year level in mind. However, each test can be used productively at two or more year levels: it is for schools to decide which test best suits any particular group of students.

The main consideration when selecting a test is to ensure its level of difficulty is appropriate to the students concerned. Tests that are too easy or too difficult will not provide precise achievement measures. Instead, there will be ceiling effects, where many students achieve perfect or near-perfect scores, and floor effects, where students answer very few questions—if any—correctly. Before using a test it is also important to check that its content is suitable for the students concerned.

Table 2 shows the year levels for which each test will be most useful. These recommendations have been based on data collected at the beginning of the school year. Schools that want to use a test near the end of the school year should bear this in mind. When in doubt, use the test recommended for the year level of the students.

Table 2 Recommended year levels for each test	
Test number	**Recommended year levels***
1	**3**, 4
2	3, **4**, 5
3	4, **5**, 6
4	5, **6**, 7
5	6, **7**, 8
6	7, **8**, 9
7	8, **9**, 10
8	9, **10**

* Bold print indicates the year level for which the test was originally developed.

3 Administration and marking

3.1 Preparation and planning

If you are administering a test, you will need to familiarise yourself with the material in advance—especially the Teacher Script. The test will probably be given only once a year, so a brief familiarisation session may be helpful for the staff who will be involved.

The instructions, texts, and questions are read to students on CD. The Teacher Script complements this, providing the test administrator with advice, instructions on what to say between test items, and the CD track numbers.

Give students the kind of notice you normally provide for regular classroom testing, without indicating that the test will be anything out of the ordinary. Where possible, administer the tests in the students' normal classroom, and preferably in the morning. Ensure the sound system to be used is working. Arrange the seating so that the students are not tempted to copy each others' answers. Make sure the room is well ventilated and that the administrator can walk around the classroom without disturbing the students.

To make the test user-friendly for younger students, Year 3 and 4 students record their answers in their test booklet; Years 5 to 10 students use an answer sheet so that their test booklets can be reused. Check that there are enough copies of the test booklets (and their matching answer sheets for Years 5 to 10) at the required level. Use the printed answer sheets, not photocopies. The answer sheets have been prepared for marking either by machine or by hand. (The quality of photocopies varies, and can adversely affect the marking process). Check that you also have the corresponding CD.

Have a supply of sharp HB or B pencils for Years 3 and 4, and some spare pens for other year levels.

Minimise disruptions during the test period. A "Do Not Disturb" notice should be hung outside the classroom door.

Table 3 can be used to help plan a suitable time allowance for tests. Younger students may require more time and help to fill out their personal details. The test time will also vary depending on the pause time taken between each track on the CD: the instructions, as well as each text and each question, are on a separate track, and there is a 5-second pause automatically programmed between each track, but some students may require longer.

Table 3 Time allowances for PAT:Listening Comprehension test	
Task	**Time**
Distribution of materials Filling in name, etc. Explanation of procedure and practice examples Test time Collection of test materials	1 min 3 min 5 min 40 min approx 1 min
Total Time	50 min approx

3.2 Administering the tests

When the students are seated and quiet, announce the name of the test and ask them to remove all materials from their desks except for a pencil (in the case of Year 3 and 4 students) or a pen (for older students). Dictionaries and thesauruses are **not** allowed. Also tell students:

- There are three types of texts in the test: texts that give information, narrative texts, and poems, and that the poems are usually short.
- The texts vary in length and are read by different readers, whose accents may reflect the texts they read.
- The term "narrator" is used in questions. Explain that "narrator" means the storyteller.

Distribute the test booklets face-up. Distribute answer sheets face-up for Year 5 to 10 students. Make sure each student is given the correct answer sheet and/or test booklet. Ask students not to complete any of their personal details until directed.

When everyone is ready, tell the students to fill in their name, gender, school, ethnicity, date, year level, and their room in the spaces provided on either the front of their test booklet for Years 3 and 4, or their answer sheet for Years 5 to 10. Explain how to fill in the spaces for their name, using one space per capital letter. Also, explain how to shade in the circles when selecting options for gender and ethnicity.

Giving instructions for the test

Ask the students to open their test booklets to the first page—the instruction page. Tell them to follow the instructions as they are read to them on the CD. Play Track 1, which incorporates the instructions as well as the example text and its questions.

When the students complete the example questions, check to see they have all found the correct place on their answer sheet or test booklet to shade their answers. If necessary, draw a set of circles on the board to demonstrate how they should be used.

Instructions to emphasise

- Students are not to look at the questions while listening to the texts on the CD—they are to look at the questions for each text **after** listening to that text and **while** they listen to the questions and the possible answers to them. It is best that students stay with the reader of the questions on the CD; i.e., tell them not to rush on. This is especially important for questions that have quotes from the text in them—students need to listen to the particular way these quotes are said.
- Students should attempt all questions, even when they are unsure of the answer. More difficult questions are usually followed by easier questions. Encourage students to keep on trying, even when they meet a harder question.
- Students should choose the **one** option they think is the **best** answer. They should not spend a lot of time shading—just enough to make their choice of circle clear.
- Students should check that they are answering each question using the matching space on the answer sheet or in their test booklet.

- If students change their minds after shading in a circle, they should carefully cross out the shaded circle and fill in the circle for the new answer they have selected, as described in the instructions. If necessary, demonstrate this on the board.

Supervising the tests

The administrator's role as supervisor or manager of the test environment is integral to the reliability of test results and to students' experience of the test. The administrator needs to judge whether students:

- need a short break during the test
- would benefit (or not) from putting their head down on their table or closing their eyes when listening to a text (If so, incorporate the instructions for this into the Teacher Script)
- need longer to answer questions (There is a 5-second gap between every track—each question is on its own track. If students need longer, press Pause. The texts are also given in the Teacher Script, so the test administrator can see the length of each text and thus foresee the transition to the questions.)
- need a gesture to start shading in their answers initially—once the test is underway students will get into the rhythm of it
- need to hear a track again if an interruption occurs.

Check that students are answering the questions in the correct place and shading the circles appropriately. You may not give the students any assistance to answer the questions. If a student experiences difficulty or appears unusually stressed, stop them and make a note of this on their answer sheet or test booklet.

After the test session

- Collect all answer sheets and/or test booklets.
- Check to see that students have given their details accurately.
- For Years 5 to 10, check that there are no markings in the booklets that could help or disadvantage a future test taker, and store the booklets for future testing.

3.3 Marking

Tests can be marked by hand. The Year 3 and 4 tests have a special marking key because these tests are presented in a different format. These keys are on pp. 61 and 63 of the Teacher Scripts for Tests 1–6 booklet. Tests for Years 5–10 have an acetate marking key that can be laid over a student's answer sheet. With the marking key positioned properly, it is easy to see which questions have been answered correctly. There is no need to write on or tick the answer sheet; simply count the number of correct answers to find the student's test score (sometimes called the "raw score"). Total the test score for Years 3 and 4 also and transfer this number to a class list, or straight onto a copy of the appropriate student report provided in Part C of this manual.

Studies have shown that hand-marked tests are often scored inaccurately. This can lead to faulty conclusions about student achievement. Some factors that contribute to errors in marking are:

- using the wrong marking acetate
- failing to notice when students have given two or more answers for the same question—**this should be marked as wrong**
- positioning the marking key wrongly, so that the questions are not aligned with their correct answers
- failing to check the answer sheet for patterns that suggest the student has struggled to use the answer sheet correctly or not taken the testing process seriously—such patterns are often masked by the marking key
- overlooking a correct answer in a run of incorrect answers (or vice versa), especially for high- or low-scoring students.

3.4 Automated marking and reporting

NZCER has developed a range of automated marking and reporting options for PAT tests. PAT:Listening Comprehension will be added to the existing suite of PAT assessments at nzcermarking.org.nz from the beginning of 2011. See www.nzcermarking.org.nz for more details, or email NZCER at marking@nzcer.org.nz.

3.5 Understanding scale scores and stanines

In itself, the number of correct answers achieved by a student on a test is difficult to interpret. It is bound to a particular test and does not show how their achievement relates to a developmental continuum that maps the whole range of knowledge and skills assessed. To address these and other shortcomings of test scores, we have constructed and described a measurement scale for the PAT:Listening Comprehension scale (the PATL scale). A test score can be converted to a location on the scale (a scale score) and then interpreted according to the scale description.

Raw test scores for the PAT:Listening Comprehension tests can be converted into locations on the PATL scale. Scale scores are expressed in units, called "patl". These scores provide a measure of students' achievement on a developmental continuum that represents increasing sophistication in listening comprehension. The conversion of a test score to a scale score takes into account the difficulty of the test, so that scaled scores on the same type of test can be compared directly, without reference to the particular level of the test or the tests the students sat.

An important property of this measurement scale is that one unit corresponds to the same amount of change in knowledge and skill—no matter where the student's achievement is located on the scale. An improvement from 10 to 20 patl units, for example, signifies the same amount of change as an improvement from 50 to 60 patl units. This equal interval property means that a student's progress in listening comprehension can be monitored over a period of years.

Every question (test item) in each of the PAT:Listening Comprehension tests has been calibrated on the listening comprehension scale. This allows a student's listening comprehension scale score to be reported in terms of the knowledge and skills required to successfully answer test items located at or below the position of the student's achievement on the scale. A student whose scale score is at the same position as a particular set of test items is expected to give correct answers for 50% of these items, and for more than 50% of the items located further down the scale.

However, no measurement can ever be totally accurate. Each scale score corresponds to the *most probable* location on the scale for a raw score on a given test, and is estimated with a margin of error that gives a range of values within which we can be fairly confident the student's achievement lies. The margin of error should be considered in all comparisons of scale scores, particularly when scores are close together.

After a test score has been converted into a scale score, it becomes possible to make comparisons with the achievement of the students in each year level of a nationally representative reference sample. The *stanine* scores for a particular year level show the position of a student's achievement in relation to that of a representative sample of students in that year level. Scale scores that fall around the average achieved by a year level in the national reference sample are assigned stanine scores from 4 to 6 (54% of students have stanines in this range). Scale scores that are well above or well below the average for a year level are placed in stanines 7–9 and 1–3, respectively.

For each test, stanines are provided at three adjacent year levels. Thus the tables given in Part C allow normative achievement to be reported in relation to three different groups of students in the national reference sample. When reporting stanines, it is important to record the year level of the reference group. For example, a Year 4 stanine score of 5 corresponds to a lower scale score than the same stanine score of 5 for Year 6.

It is also important to remember that the data used in the normative reporting were collected at the beginning of a school year. If students are assessed towards the end of the year, it might be more appropriate to report stanines with reference to the group at the next year level up.

Part B of this manual gives more information about PAT:Listening Comprehension scale scores and stanines and how they were derived.

3.6 Converting test scores to scale scores

Test score conversion tables are used to convert test scores on each PAT:Listening Comprehension test into scale scores, and then into their associated stanines at selected year levels. Part C of this manual provides a table for each test. It is important to check that you are using the correct table.

The sample test score conversion table shown in Figure 1 (p. 21) gives the results of a fictitious student, Ben Walker. He has completed PAT:Listening Comprehension Test 4 and answered 19 questions correctly. His teacher has circled test score 19, scale score 50.6 patl units with its associated error range of 3.5 patl units, and the Year 6 stanine level 4. The teacher could have chosen stanines for Year 5 or Year 7 instead, depending on the time of year of the testing or the purpose of the student report.

The table shows the measurement error associated with each scale score, with the smallest errors occurring near the mid-test score. Ben's most probable scale score has an error of 3.5 patl units, meaning that we can be reasonably confident that his achievement level is somewhere between 3.5 patl units below and above his score of 50.6 ; that is, in the range 47.1–54.1 patl units.

3.7 Age-level normative information

Age-level stanines are not provided for PAT:Listening Comprehension. As part of the norming study, data on student ages were collected and a study of achievement at different age levels was made. Across each year group tested no consistent correlation was found between achievement on the test and the ages of the students. In other words, younger and older students within each year group did equally well. This suggests that it is the year group students are in—not their age *per se*—that explains more of the variation in achievement. More information on this study can be found in Part B of this manual.

4 Reporting the test results

4.1 The individual student report

Test results for individual students can be reported using the PAT:Listening Comprehension student reports, which create a map of each student's achievement over the whole test. Part C provides templates of the PAT:Listening Comprehension student report for each test. (A sample report is also shown in Figure 2.) Figure 3 explains the different sections of the student report.

To construct a student report

1. Make a copy of the appropriate test report template (given in Part C of this manual).
2. Fill in the student's name and year level.
3. On the score conversion table (see Figure 1) for that particular test, find the student's test score, scale score and the range of error. Transfer these figures to the box at the top right of the report.
4. Locate the test score on the scale and circle it. Draw a horizontal line across the page at this position.
5. Draw broken horizontal lines above and below this line to show the error range for the scale score.
6. Circle each question that has been answered correctly.
7. Underline each question that has not been answered.

In the report shown in Figure 2, these steps have been completed for our fictitious student, Ben Walker. The report shows that Ben attempted every question and answered 19 of them correctly. Note how the horizontal line crosses his test score (19), the PAT:Listening Comprehension scale score (50.6 patl), and the stanines corresponding to three different year levels (stanine 5 at Year 5, stanine 4 at Year 6, and stanine 4 at Year 7). The measurement error is shown by the dashed lines, which are 3.5 patl above and 3.5 patl below his scale score. We can be about 70% confident that Ben's true location on the scale lies somewhere in this interval.

The report also shows how Ben performed on the individual questions for each text. Different symbols are used to show the different question types: retrieval, local inference and global inference. Ben's scale score can be compared with the location on the scale of the different questions.

For example, question 7 in 'Sione Wakes Up' is positioned well above his location on the scale. Given this difference he would not be expected to get it correct. Question 1 in 'Pass It On', on the other hand, is well below his scale score, so it would be a surprise if he got it wrong. Finally, questions that lie close to his scale score are ones he has an even chance of answering correctly.

Ben's pattern of responses is fairly typical for a student with a test score of 19 on Test 4. Most of the questions he has answered correctly are at or below the level of his scale score, while most of the incorrectly answered questions are above that level.

4.2 The scale descriptions

Ben's diagnostic report should be read in conjunction with the scale descriptions for listening comprehension given in Part C of this manual. Each description describes the types of listening comprehension knowledge and skills required to answer test items located at different parts of the scale.

The listening comprehension scale descriptions cover eight pages; the first page provides an introduction for the seven pages of descriptions that follow. Each of these examines a particular type of competency; for instance, "using abstract information" and "using implied information". We can use the scale descriptions to get a sense of the skills and knowledge students are developing as they progress along the listening comprehension scale.

Figure 1 Sample score conversion table for PAT:Listening Comprehension Test 4, showing results for Ben Walker

LISTENING TEST 4

Score Conversion Table: **Test 4**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 5 stanine	Year 6 stanine	Year 7 stanine
35	96.1	>10.2	9	9	9
34	85.8	10.2	9	9	9
33	78.5	7.4	9	9	9
32	74.0	6.1	9	8	8
31	70.7	5.4	8	8	7
30	68.0	5.0	8	7	7
29	65.7	4.6	7	7	7
28	63.7	4.4	7	7	6
27	61.9	4.2	7	6	6
26	60.2	4.0	6	6	6
25	58.7	3.9	6	6	5
24	57.2	3.8	6	5	5
23	55.8	3.7	6	5	5
22	54.5	3.6	5	5	5
21	53.1	3.6	5	5	4
20	51.9	3.6	5	5	4
19	50.6	3.5	5	4	4
18	49.4	3.5	5	4	4
17	48.1	3.5	4	4	4
16	46.9	3.5	4	4	3
15	45.6	3.6	4	3	3
14	44.3	3.6	4	3	3
13	43.0	3.6	3	3	3
12	41.7	3.7	3	3	2
11	40.3	3.8	3	3	2
10	38.8	3.9	3	2	2
9	37.3	4.0	2	2	2
8	35.7	4.1	2	2	1
7	33.9	4.3	2	1	1
6	31.9	4.6	1	1	1
5	29.6	4.9	1	1	1
4	27.0	5.4	1	1	1
3	23.7	6.1	1	1	1
2	19.3	7.3	1	1	1
1	12.0	10.2	1	1	1
0	1.8	>10.2	1	1	1

Figure 2 Individual student report for Ben Walker

LISTENING TEST 4

PAT: Listening Comprehension Student Report

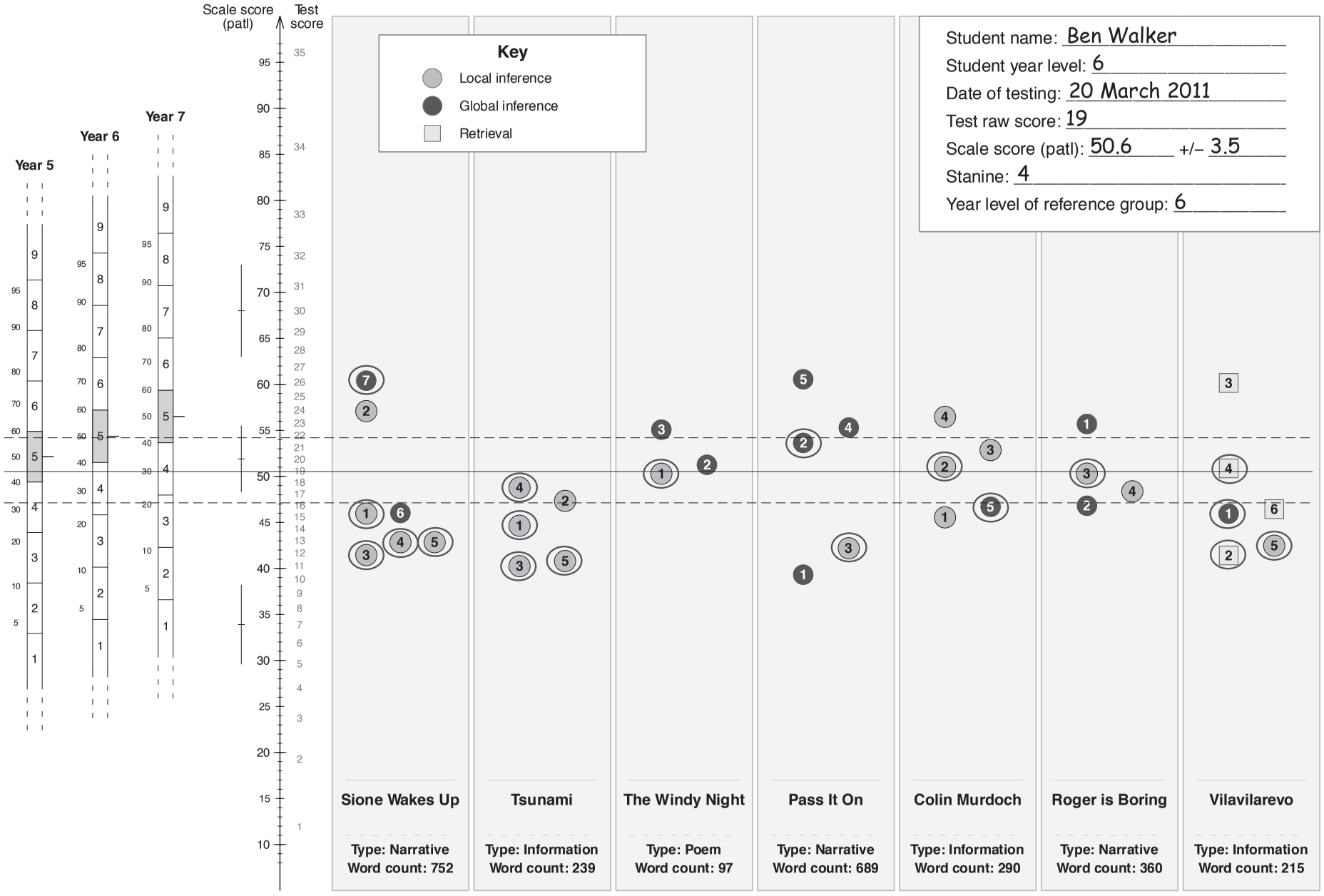
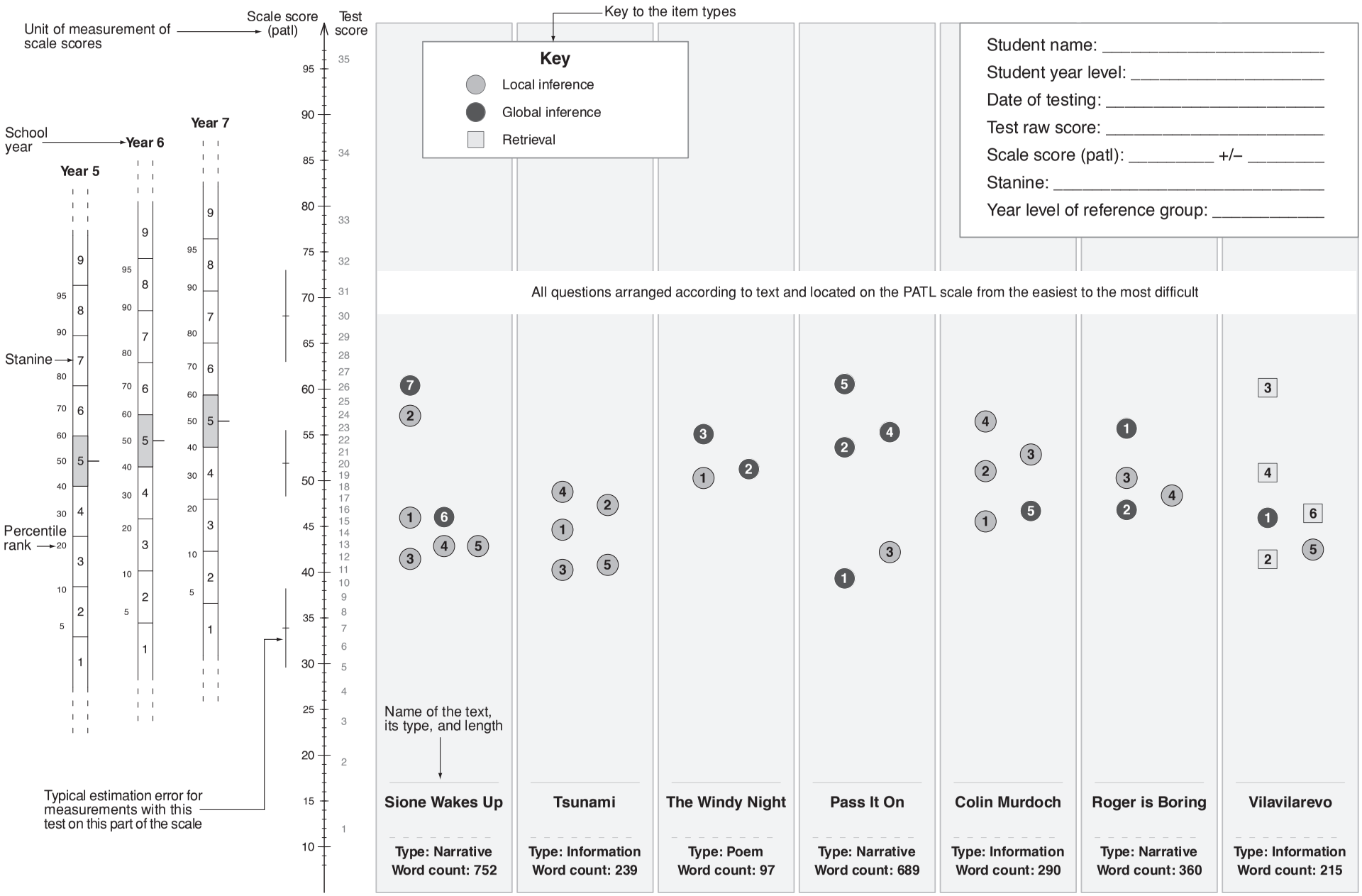


Figure 3 Key to the individual student report

PAT: Listening Comprehension Student Report



5 Using the test results

When used in conjunction with teacher observations and assessment data from a range of other sources, information from PAT:Listening Comprehension can help inform decisions about teaching and learning. In particular, the student reports and scale descriptions can be used to outline the kinds of listening comprehension competencies students are developing and how their progress compares with national norms. The following sections provide some guidance on how the test results might be used.

5.1 Measuring progress over time

- The PAT:Listening Comprehension scale provides an excellent tool for measuring progress over time. Scale scores can be plotted and trajectories monitored to provide a profile of growth for groups and individuals. Figures 7 (page 33) can be used to compare a student's or group's progress with national norms for listening comprehension.
- Generally, scale scores for an individual student should increase from year to year. If they do not increase, further investigation is warranted. Note, however, that due to measurement error the actual observed gain in the scale score for an individual is relatively imprecise, so it is unwise to make definitive statements about exactly how much an individual has improved. This lack of precision can be exacerbated when the time period between measurements is less than a year. For a group (such as a class, or year level), the average gain in scale score over time is a more reliable measure and can be used to make meaningful statements about the amount of change in group performance.
- It is important to note that listening progress over the year levels appears to move more slowly than reading progress as measured by PAT:Reading Comprehension (Darr et al., 2008). On average, students move up the PAT:Listening Comprehension scale at the equivalent of half a stanine each year, whereas average yearly growth on the PAT:Reading Comprehension scale is about one stanine. This finding is also evident in National Education Monitoring Project data (Crooks & Flockton, 2005; Crooks, Flockton, Smith, & Smith, 2007; Flockton & Crooks, 2003). Relatively slower listening progress may occur because prereaders, and those still heavily focusing on decoding print, make extensive use of listening as a means of comprehension, while in later years the printed word tends to increasingly dominate, at least in terms of classroom instruction.

5.2 Grouping students with similar needs and competencies

Scale scores can be used to help arrange instructional groups by achievement level. This is true even if the students involved take tests at different levels, because the scale scores locate student achievement on a common scale. However, any initial grouping decisions based on PAT:Listening Comprehension results should be reviewed in response to ongoing classroom-based assessment.

5.3 Identifying readers requiring support

Stanines should be used by teachers to help identify students whose scale scores indicate low achievement for their year level. Students who score stanine 3 and below for their year level certainly require closer attention. As noted above, other sources—such as the teacher's knowledge of the child and previous records—may alter the priorities for further investigation.

5.4 Selecting able students to be considered for enrichment programmes

Stanines can also help teachers identify students whose scale scores indicate high achievement for their class level. Again, this should be considered alongside other sources of information before students are selected for enrichment work.

5.5 Locating areas of strength and weakness within a class

Standardised tests such as PAT:Listening Comprehension can be used to identify areas of overall weakness and strength in classroom or school programmes. Discrepancies between achievement in different subject areas can be examined; for instance, between reading comprehension and listening comprehension. When comparing achievement on tests in different subject areas, year-level stanines rather than scale scores should be used.

5.6 Broad diagnostic uses

Used with care, PAT:Listening Comprehension can provide a broad level of diagnostic information. By completing the student reports and reading the scale descriptions, teachers can be provided with information about the kinds of listening comprehension skills that students are having success with. It is possible, for instance, that some students may do well on test items requiring local inferences but show serious weakness when making global inferences.

In general, however, teachers should take care not to reach definite conclusions about specific strengths and weaknesses based on the results from only one test. Rather, PAT:Listening Comprehension should be seen as a starting point, where answer patterns (such as unexpected incorrect or correct answers) may reveal clues that can lead to further investigation.

Performance on different texts

Students' prior knowledge and experiences, along with their engagement with certain text types or certain topics, affects their ability to listen and comprehend. Investigation of the student report for listening comprehension might show that a student has had more or less success on a particular text or text type than is expected given the student's location on the scale.

Performance on different items or item types

The student report for listening comprehension can be used to identify how well individual students have performed on the different types of test items. A pattern of incorrect answers on items that require students to make global inferences, for instance, might indicate the need for specific instruction in this skill. When looking for patterns, the difficulty (scale locations) of the items should be taken into account. For example, if the global inference questions in a particular test are consistently more difficult (located higher on the scale) than the questions assessing local inference, then the student will be expected to answer more of the local inference questions correctly.

5.7 Communicating test results to students and parents/caregivers

Test results need to be communicated to students and parents/caregivers with tact and skill. We cannot give a set of rules for this—teachers must use their professional judgement and experience to tailor this process to the individual student and his or her parents—but we can offer some general guidelines.

Both students and parents need to feel that the testing process is about the students' welfare and development. Take care to stress the positive aspects of test performance and to provide encouragement. Sensitivity is important, because low test scores can lead to students becoming discouraged, particularly if scores are presented as a fixed measure of innate ability. A more positive way to deal with low scores is to see them as an indication of present limitations that can, with work, be turned around.

The PAT:Listening Comprehension scale score can be an encouraging indicator because it generally grows from year to year as a student's understanding and maturity increase. Stanines, on the other hand, tend to remain static. They reflect the position of any one student relative to students in the national reference group, and will only increase if the student is advancing more rapidly than the reference group.

When talking to younger students, it is often best to describe their results in general terms, such as "like most others". This can be supplemented with more specific information about particular strengths and weaknesses. Older students may benefit from a more detailed interpretation that includes and explains scale scores, stanines and the student reports.

When reporting scores to parents or caregivers, include an explanation of what the scores mean and the margin of error for the measurements. The key to the student report (Figure 3, see above) can be used to help parents better understand how their child is progressing. It is important to remember that a PAT:Listening Comprehension score is only one way of assessing a student's achievement, and other indicators of progress are also needed to give a full and accurate picture.

PART B

Technical Information

Part B of this *Teacher Manual* provides technical information about the revision of PAT:Listening Comprehension, including the construction of a Rasch scale to describe students' progress in listening comprehension.

Part B is divided as follows.

- Section 6 describes the construction of the scale, along with a technical explanation of the scale properties.
- Section 7 summarises the properties of the PAT:Listening Comprehension scale.
- Section 8 gives a brief overview of the revision of the tests.
- Section 9 is devoted to the norming study describing the sample, its properties and the process of estimating national norms.
- Section 10 discusses the reliability and precision of estimates, and validity.
- Section 11 describes properties of the tests.
- Section 12 summarises further analysis on two demographic sub-groups and provides some indicative correlations with PAT:Reading Comprehension.

6 Construction of the scales

This graduated series of PAT:Listening Comprehension tests has been designed to address the needs of students in Years 3 to 10. A student's achievement in a subject area assessed by a graduated series of tests such as this has to be expressed in the same metric if progress is to be measured meaningfully across several year levels. In other words, the results from each test in the series must be able to be mapped to the same underlying scale.

In education, creating such a common scale is achieved by applying the Rasch measurement (RM) model, which was developed by the Danish mathematician Georg Rasch in the 1950s (Rasch, 1980). RM has been applied to the construction of interval scales in educational and other fields for over 40 years. NZCER's Progressive Achievement Tests in Mathematics and Reading Comprehension and Vocabulary have already undergone major revisions (in 2006 and 2008, respectively) using this methodology.

In this project the statistical programming package R (Ihaka,1996) was used for item calibration and all other statistical analysis. R was also used to create most of the graphics in this manual.

6.1 Measuring progress on a described scale

Development in listening comprehension can be thought of as a journey that can be mapped along a continuum. As students progress along the continuum, their knowledge and skill increase and become more sophisticated. How much knowledge and skill in listening comprehension a student has is not directly observable, but can be inferred from their responses to test items designed to probe the competencies involved. Each test item will require a certain level of competence in order to be answered correctly.

We can measure progress in listening comprehension by constructing a scale that represents this continuum. Such a scale allows us to locate the achievement levels of different students in listening comprehension. It also allows us to locate the levels of listening comprehension knowledge and skill required to correctly

answer different test items. Each student achievement in listening comprehension has a fixed location on a scale (student location), and each item (in terms of its difficulty in relation to other items) has a fixed location on the *same* scale (item location).

A student located on the scale at the same place as a test item will have a 50% chance of answering that item correctly. Items located higher on the scale than the student will be more difficult for that student to answer correctly, and items located lower on the scale than the student will be easier to answer correctly.

In traditional test theory, test scores (sometimes called raw scores) are used to indicate levels of student achievement. However, test scores depend as much on the complexity of the items in the test as on the skill levels of the students. In a similar way, the facility of each test item (usually reported as the percentage of correct responses) depends as much on the level of skill of the particular group of students taking the test as on the level of skill that each item requires.

As a result, test scores and item facilities must always be interpreted in terms of the particular test used to generate the test score and the particular sample of students who sat the test containing the items. There is no independent scale of measurement in traditional test theory that defines a continuum on which both student achievement and item difficulty can be located.

Separating the estimation of item locations from student achievement is central to objective measurement. Objective measurement requires the relative location of items to be the same for students placed at any location on the scale, and it requires the location of achievement to be independent of the particular set of items included in the test or tests administered. Both requirements are identical to those used for measurement in the physical sciences. When height is measured, for example, everybody is expected to agree on which of two objects is the taller, and that the results do not depend on which rulers are used.

6.2 Constructing a Rasch measurement scale

Rasch Measurement (RM) assumes that the observed achievements of students and the observed relative difficulty of test items can be represented by fixed locations on the same interval scale. Each location on the scale is said to correspond to a certain amount of the attribute being measured. RM then proposes a mathematical model to predict the probability of success for any student on any of the items calibrated onto this scale.

According to the model, this probability depends only on the difference between the respective scale locations of student and item. A student who is at the same scale location as a group of items is expected to answer 50% of these items correctly. The same student is expected to answer correctly more than 50% of items located lower on the scale, and fewer than 50% of items located higher on the scale. A consequence of this assumption is that the location of items on the scale is independent of the distribution and position of student locations on the scale. The same item locations are expected, within a margin of error and in the absence of misfit, for different distributions of student achievement.

The construction of an RM scale—such as a listening comprehension scale—involves writing items designed to assess listening skills, trialling the items, and collecting norming data with test forms assembled using trialled items. By including common items in trial test forms, or administering more than one test to the same selection of students, all items can be linked across the different test forms. The data are then analysed with the Rasch model, initially to estimate the relative location of items on the scale (item calibration). A meticulous fit analysis, examining each item in the light of both statistical and graphical fit indicators, shows how well the data fit the measurement model and exposes any items that did not perform as well as expected.

Once the scale is finalised, it becomes possible to estimate student locations on the scale and obtain the distribution of student achievement by year level. During the development of PAT:Listening Comprehension, RM was applied in the piloting, trialling and norming stages to make sure all items fitted their respective

scales within the required tolerances. Items that did not fit the model satisfactorily were excluded from the final tests.

The practical outcome of applying RM is a bank of items located on a single measurement scale that can be described to show the range of knowledge and skill levels in listening appropriate to students at different achievement levels. Items from this bank can then be combined into test forms that best target the location of different groups of students (e.g., students at different year levels).

The ability of RM to transform raw test scores into scale scores allows us to measure achievement without having to indicate the level of difficulty of the test administered to a student. It is important to note, however, that not all the tests are suitable for administration to any one group of students. To be suitable, the test has to match their range of knowledge and skills. A test that is relatively very easy or very difficult will result in large errors of measurement on the scale.

Once constructed, a scale can be described qualitatively by examining the items in each region of the scale and summarising the knowledge and skills that are characteristic of these items. A student located at any point on the scale is likely to have mastered the skills below that location and is less likely to have mastered those above. A student's achievement in PAT:Listening Comprehension can be reported in terms of both the knowledge and skills exhibited (formative reporting) and the relative position in the distribution of locations of students in a given year level (normative reporting).

6.3 The PAT:Listening Comprehension scale

Figure 4 shows the test items used in the construction of the PAT:Listening Comprehension scale (PATL scale), by test, and Figure 5 shows test items for the comprehension tests grouped by text. Items were calibrated onto the RM scale using data from the norming study.

The scale reports achievement in terms of a special PAT:Listening Comprehension measurement unit called a "patl". The PATL scale in Figure 4 shows achievement from 0 patl to 105 patl. Item difficulty ranges from around 30 patl (items 21 in Test 1, 24 in Test 2 and 17 in Test 3) to around 85 patl (items 11, 16, 19 and 27 in Test 8).

6.4 Item numbering

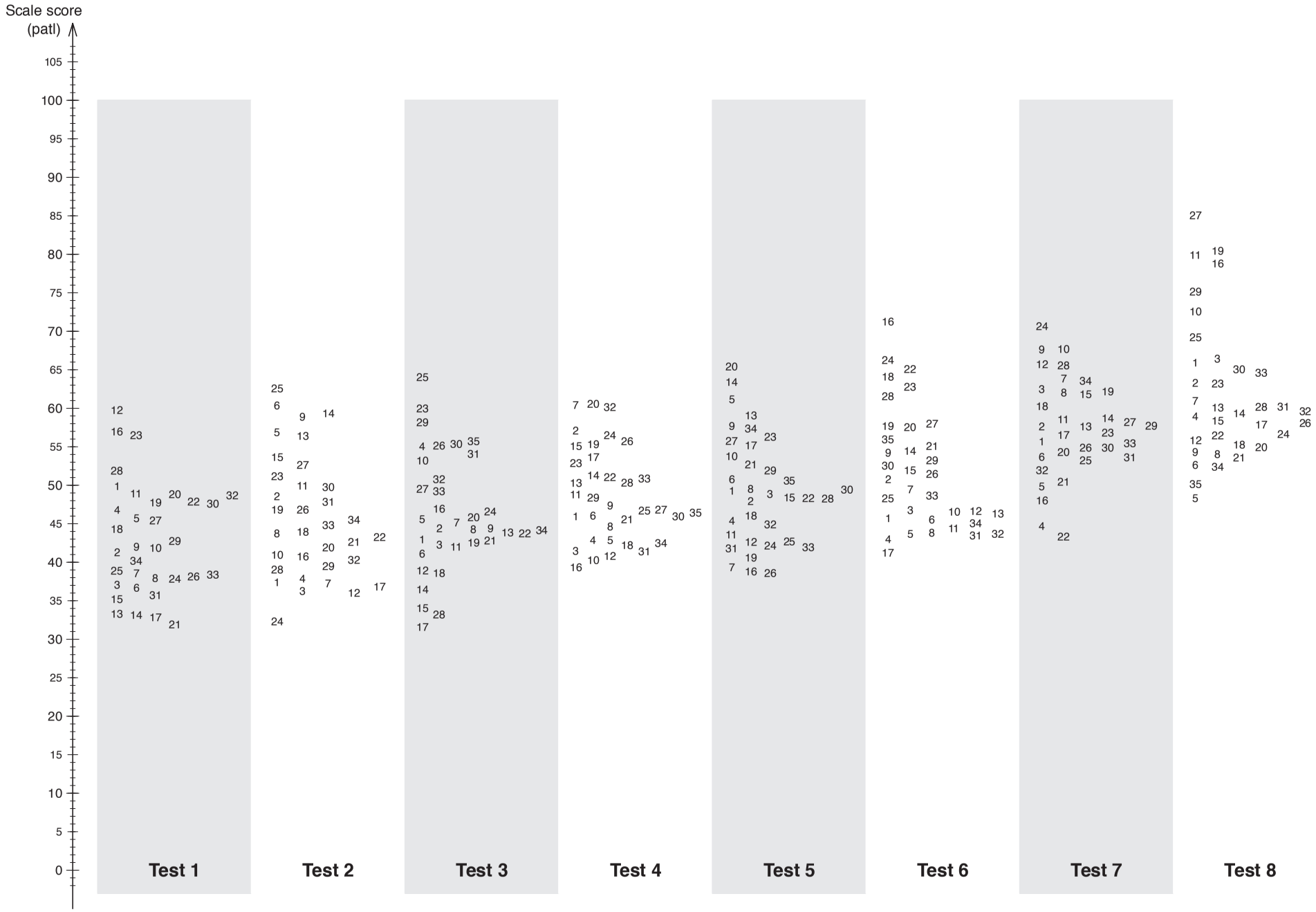
In an effort to make the published test booklets and answer sheets less confusing for students, text titles have been printed and then questions numbered starting from 1 within each text. However, in some graphics it is more convenient or logical to number the items across the whole test. For example, in the "items by test" graphic (Figure 4) the items in Test 1 are numbered 1 to 34, whereas in the "items by text" graphic (Figure 5) items for Test 1 are numbered starting at 1 for each text. A cross-referencing key is provided in Figure 6.

Suppose we are interested in following up the questions from the text 'Big Birds' in Test 3. In Figure 5 the items are numbered 1–6 within the text. Figure 6 tells us that these items relate to numbers 22–27 in Test 3. Depending on which graphic or table is of interest, the appropriate numbering system can now be used. Alternatively, suppose we are interested in the most difficult few questions in Test 4. We can find these from Figure 4: they are items 7, 20 and 32. Referring to Figure 6, we can tell that these items belong to 'Sione Wakes Up', 'Pass It On' and 'Vilavilarevo', respectively.

6.5 The achievement of the norming sample, by year level

Figure 7 shows the distribution of student achievement, by year level, on the PATL scale. As can be seen, while there is substantial overlap between year levels, the mean scale score for each year level increases gradually at a fairly constant rate.

Figure 4 Listening comprehension test items grouped by test



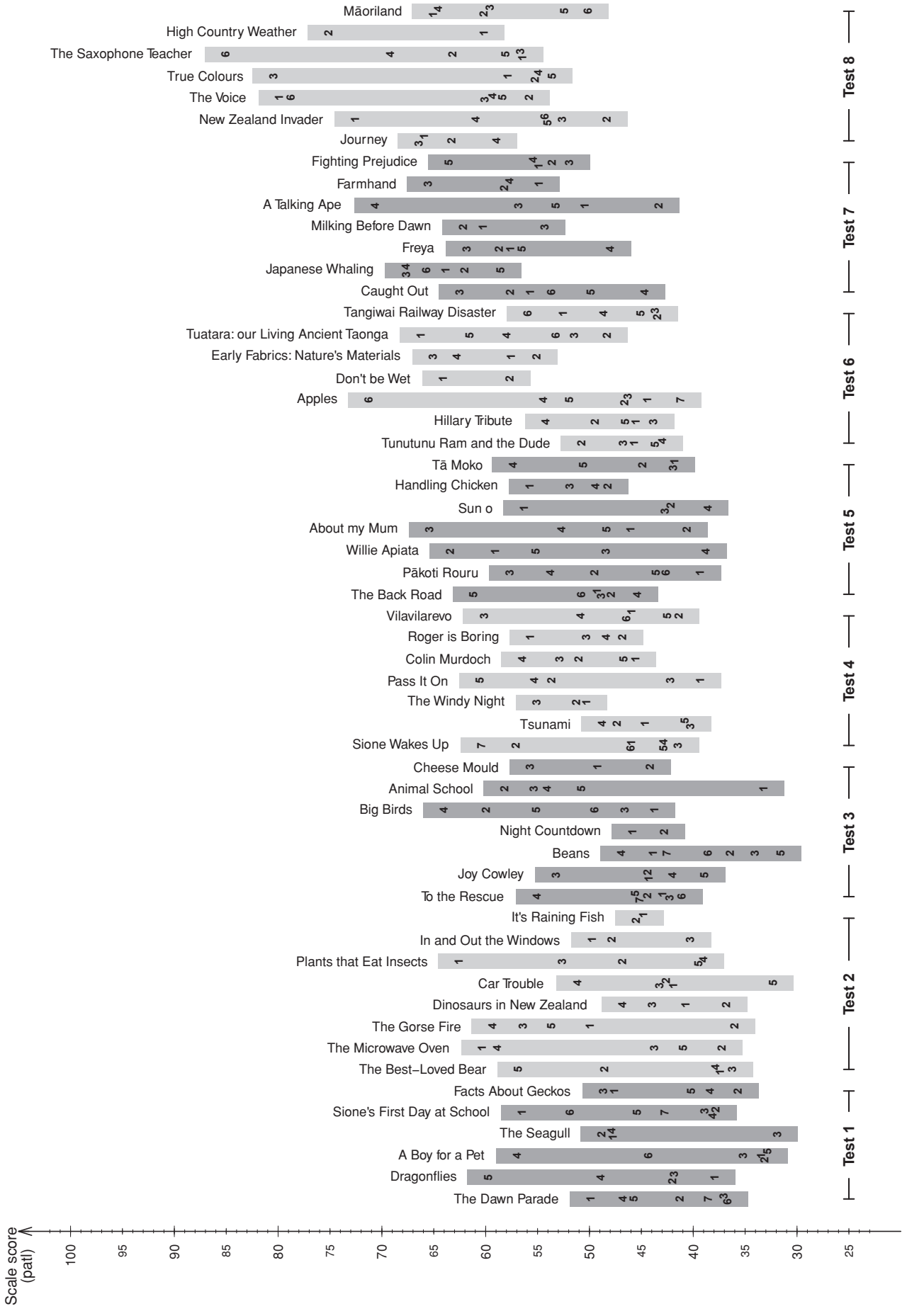


Figure 5 Listening comprehension items grouped by text

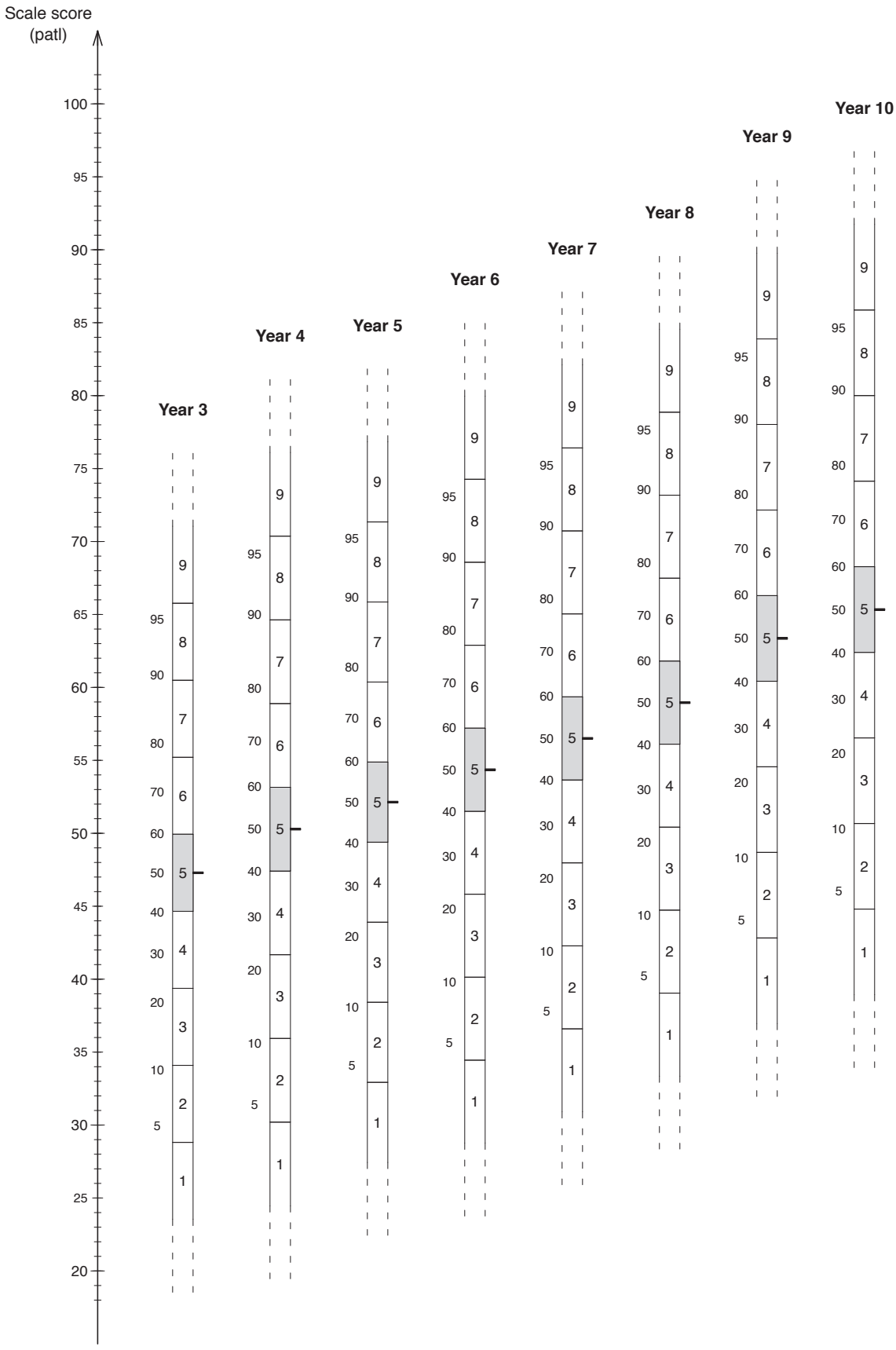
Figure 6 Item number cross-reference table

Test 1								Test 2								Test 3							
The Dawn Parade	Dragonflies	A Boy for a Pet	The Seagull	Sione's First Day at School	Facts About Geckos			The Best-Loved Bear	The Microwave Oven	The Gorse Fire	Dinosaurs in New Zealand	Car Trouble	Plants that Eat Insects	In and Out the Windows	Raining Fish	To the Rescue	Joy Cowley	Beans	Night C'down	Big Birds	Animal School	Cheese Mould	
1 – 1	1 – 8	1 – 13	1 – 19	1 – 23	1 – 30			1 – 1	1 – 6	1 – 11	1 – 16	1 – 20	1 – 25	1 – 30	1 – 33	1 – 1	1 – 8	1 – 13	1 – 20	1 – 22	1 – 28	1 – 33	
2 – 2	2 – 9	2 – 14	2 – 20	2 – 24	2 – 31			2 – 2	2 – 7	2 – 12	2 – 17	2 – 21	2 – 26	2 – 31	2 – 34	2 – 2	2 – 9	2 – 14	2 – 21	2 – 23	2 – 29	2 – 34	
3 – 3	3 – 10	3 – 15	3 – 21	3 – 25	3 – 32			3 – 3	3 – 8	3 – 13	3 – 18	3 – 22	3 – 27	3 – 32		3 – 3	3 – 10	3 – 15		3 – 24	3 – 30	3 – 35	
4 – 4	4 – 11	4 – 16	4 – 22	4 – 26	4 – 33			4 – 4	4 – 9	4 – 14	4 – 19	4 – 23	4 – 28			4 – 4	4 – 11	4 – 16		4 – 25	4 – 31		
5 – 5	5 – 12	5 – 17		5 – 27	5 – 34			5 – 5	5 – 10	5 – 15		5 – 24	5 – 29			5 – 5	5 – 12	5 – 17		5 – 26	5 – 32		
6 – 6		6 – 18		6 – 28				6 – 6								6 – 6		6 – 18		6 – 27			
7 – 7				7 – 29				7 – 7								7 – 7		7 – 19					

Test 4								Test 5								Test 6							
Sione Wakes Up	Tsunami	Windy Night	Pass It On	Colin Murdoch	Roger is Boring	Vilavilarevo		The Back Road	Pākoti Rouru	Willie Apiata	About my Mum	Sun o	Handling Chicken	Tā Moko		Tunutunu Ram and the Dude	Hillary Tribute	Apples	Don't be Wet	Early Fabrics	Tuatara: our Living Ancient Taonga	Tangiwai Railway Disaster	
1 – 1	1 – 8	1 – 13	1 – 16	1 – 21	1 – 26	1 – 30		1 – 1	1 – 7	1 – 13	1 – 18	1 – 23	1 – 27	1 – 31		1 – 1	1 – 6	1 – 11	1 – 18	1 – 20	1 – 24	1 – 30	
2 – 2	2 – 9	2 – 14	2 – 17	2 – 22	2 – 27	2 – 31		2 – 2	2 – 8	2 – 14	2 – 19	2 – 24	2 – 28	2 – 32		2 – 2	2 – 7	2 – 12	2 – 19	2 – 21	2 – 25	2 – 31	
3 – 3	3 – 10	3 – 15	3 – 18	3 – 23	3 – 28	3 – 32		3 – 3	3 – 9	3 – 15	3 – 20	3 – 25	3 – 29	3 – 33		3 – 3	3 – 8	3 – 13		3 – 22	3 – 26	3 – 32	
4 – 4	4 – 11		4 – 19	4 – 24	4 – 29	4 – 33		4 – 4	4 – 10		4 – 21	4 – 26	4 – 30	4 – 34		4 – 4	4 – 9	4 – 14		4 – 23	4 – 27	4 – 33	
5 – 5	5 – 12		5 – 20	5 – 25		5 – 34		5 – 5	5 – 11		5 – 22			5 – 35		5 – 5	5 – 10				5 – 28	5 – 34	
6 – 6						6 – 35		6 – 6	6 – 12							6 – 6					6 – 29	6 – 35	
7 – 7								7 – 7															

Test 7								Test 8							
Caught Out	Japanese Whaling	Freya	Milking Before Dawn	A Talking Ape	Farmhand	Fighting Prejudice		Journey	New Zealand Invader	The Voice	True Colours	The Saxophone Teacher	High C'try Weather	Māoriland	
1 – 1	1 – 7	1 – 13	1 – 18	1 – 21	1 – 26	1 – 30		1 – 1	1 – 5	1 – 11	1 – 17	1 – 22	1 – 28	1 – 30	
2 – 2	2 – 8	2 – 14	2 – 19	2 – 22	2 – 27	2 – 31		2 – 2	2 – 6	2 – 12	2 – 18	2 – 23	2 – 29	2 – 31	
3 – 3	3 – 9	3 – 15	3 – 20	3 – 23	3 – 28	3 – 32		3 – 3	3 – 7	3 – 13	3 – 19	3 – 24		3 – 32	
4 – 4	4 – 10	4 – 16		4 – 24	4 – 29	4 – 33		4 – 4	4 – 8	4 – 14	4 – 20	4 – 25		4 – 33	
5 – 5	5 – 11	5 – 17		5 – 25		5 – 34			5 – 9	5 – 15	5 – 21	5 – 26		5 – 34	
6 – 6	6 – 12								6 – 10	6 – 16		6 – 27		6 – 35	

Figure 7 Distribution of student achievement in listening comprehension by year level



7 Properties of the PAT:Listening Comprehension scales

RM produces measures that are recorded on an interval scale. This means, for example, that an increase of one patl unit in a particular part of the PATL scale represents an equivalent one patl unit increase in knowledge and skills anywhere else on the scale. Test scores and percentiles do not have this property—they are ranks rather than measures. Each change in a test score or percentile represents a different amount of change in the knowledge and skills represented by the scale score. For students achieving around the mean score in a test, a change of one mark or one percentile represents only a very small change in the scale score. However, for students achieving towards the top or bottom of their year level, a change of one percentile or one test mark represents a much larger change in the scale score.

As with other commonly used scales (e.g., length, temperature or weight), the choice of numerical values and the names of units to assign to an RM scale is arbitrary. The unit of measurement used to express locations on the PAT:Listening Comprehension scale (patl) results from the transformation of the unit used by the Rasch model to estimate locations on the scale. This unit is called the logit. A logit is not a unit of fixed size for all Rasch scales: its size is dependent on the particular scale being constructed.

For the PATL scale, the transformation from logits to patl is:

PAT:Listening Comprehension scale score (patl) = 10 × logit + 50

It follows that:

- 1 patl unit is equivalent to 0.1 logits
- the mean location of all the listening comprehension items calibrated onto the listening scale is set at 50 patl
- a student located at any point on a scale *(b* patl units, for example) is expected to answer correctly 50% of the items at that location, around 30% of the items 10 patl units higher, and around 10% of the items 20 patl units higher; similarly, the same student located at *b* patl is expected to answer correctly just over 70% of the items 10 patl units lower, and about 90% of the items located 20 patl units lower than *b*.

8 Construction and revision of the PAT:Listening Comprehension tests

8.1 Revising the tests

The review process began with specifications being drawn up for the new tests that described the structure of the tests and their content. This was used as a basis for writing new items. A panel of comprehension experts reviewed a representative sample of the items and gave advice on the proposed aims of the review. A panel of teachers reviewed all of the items. Items were piloted with small numbers of students, and this information was used to help select items for the much larger trials.

8.2 The national trial

Fifteen trial forms were constructed for the national trial. Each of the tests included a number of items common to another test form. Linking items across tests and year levels in this way allowed all of the item difficulties from the tests to be located on a single scale. The 15 trial forms were used in a national trial with 1,745 students from Years 3 to 10, who were selected using a stratified random sampling design. The trial was designed to ensure that each of the items had good Rasch measurement properties and suited the year levels to which it had been allocated.

8.3 The norming forms

Data from the national trial, together with comments from teachers who administered the trial tests to their classes, were used to help construct eight new forms for the norming study. Each of the norming

forms was designed to suit a specific year level. Seven additional forms were constructed to provide item links between the eight norm forms. A total of 15 forms were therefore prepared for the norming study. Details of the norming study can be found in Section 9

8.4 The published tests

The best performing items from the norming study that targeted the range of achievement found at each of Years 3 to 10 were combined into the eight final published tests. Properties of the published tests are provided in Section 11.

9 The norming study

9.1 The norming sample

The calibration of an RM scale is most precise when the analysis is based on a large number of students' responses. Although the sample is expected to reflect the general characteristics of the population, RM does not require it to be representative in the statistical sense. The calibration of the items is independent of the shape and location of the distribution of student achievement on the scale.

On the other hand, to report achievement in terms of meaningful norms, it is essential to use a sample that is representative of the national population. A stratified random sample of schools ensured proper representation of the national population at all year levels. Stanines for PAT:Listening Comprehension were calculated for each year level sampled. (See Figure 8 on page 38 below for a detailed explanation of stanines.)

Stratification of the sample

The stratification variables for the national norming sample were school decile and school size. Deciles were grouped 1–3, 4–7 and 8–10, and schools were assigned to "small", "medium" or "large" depending on the roll size of the year levels of involved. The sample was selected in four parts: a separate sample for Years 3–4, Years 5–6, Years 7–8 and Years 9–10.

Recent experience at NZCER has shown that most of the variance in achievement on similar tests (e.g., Progressive Achievement Tests in Mathematics and Reading Comprehension) lies within a school rather than between schools. With this in mind, it was decided to select complete year levels from some schools and a selection of classes (one to three classes) from other schools. Selecting a complete year level ensures we have complete representation of the range of ability. Selecting a small number of classes from other schools helps to ensure representation across a range of schools.

Once the sample was selected, it was carefully inspected to ensure reasonable representation across other variables such as school type, school location and ethnic proportions. Non-response and refusals were managed by drawing up a randomised replacement list within each stratum. If a school chose not to participate in the project, a replacement school was selected from the list.

One hundred and thirteen schools participated in the norming study, most schools providing a selection of students at two year levels. Tables 4 and 5 show the number of schools and students in the norming sample, respectively.

Table 4 Number of schools in the national sample by year level and decile group

	Number of schools							
Decile	**Year 3**	**Year 4**	**Year 5**	**Year 6**	**Year 7**	**Year 8**	**Year 9**	**Year 10**
Low	6	6	7	7	9	9	6	6
Mid	13	13	15	15	9	9	9	9
High	14	14	11	11	10	11	5	4
Total	33	33	33	33	28	29	20	19

Table 5 Number of students in the national sample by year level and decile group

	Number of students								
Decile	**Year 3**	**Year 4**	**Year 5**	**Year 6**	**Year 7**	**Year 8**	**Year 9**	**Year 10**	**Total**
Low	416	406	398	362	552	573	274	268	3,249
Mid	549	513	703	732	538	569	563	540	4,707
High	695	704	741	672	589	607	282	223	4,513
Total	1,660	1,623	1,842	1,766	1,679	1,749	1,119	1,031	12,469

The achieved sample

Participating schools were asked to select groups of students at the required year levels to represent their school in the PAT:Listening Comprehension tests. It was requested that student groups reflect the full range of abilities at each year level. Where schools were not providing complete year levels, large schools were asked for three class groups, middle-sized schools for two class groups, and smaller schools for just one class. Complete year levels were provided by 34 schools, which were spread evenly across the 12 strata.

The trial forms were sent to schools in mid-March 2009. Schools were given until the end of the month to complete the test.

Once the test forms had been returned, data from the entire sample were used to calibrate all the items developed for PAT:Listening Comprehension onto a scale. A careful analysis of item performance was carried out. Items that failed to fit the Rasch model were removed for the final construction of the scales.

During the analysis of the sample a few unexpected anomalies were discovered. A pragmatic approach was taken to rectify these issues, as follows.

- The achieved Years 9 and 10 sample was smaller than expected. A decision was taken to establish norms for these year groups using the numbers available.
- A few schools had to be eliminated from the norming study because their results did not appear to represent the school as a whole. However, these schools' data were able to be used in the calibration process.
- The standard deviations estimated from the sample for Years 8 and 9 showed quite a marked deviation from the estimated standard deviations at other year levels. It was decided to adjust the standard deviations for Years 8 and 9 to reflect those of adjacent year groups.

Post-stratification of the sample

If a sample is a slightly inaccurate model of the population of interest, it does not affect the calibration of items. Sample bias can, however, affect the estimates of year-level norms. Consequently, the norming study included a post-stratification procedure to remove sample bias. The variables considered in post-stratification were decile (grouped) and gender.

The achieved sample showed some deviation from national proportions of students in some strata. An iterative subsampling procedure was employed to see if any adjustments should be made to the initial estimates of national norms. The procedure involved repeatedly subsampling (500 times) the achieved sample randomly according to national proportions with respect to decile group and gender. Each subsample consisted of approximately 70% of the original sample. The estimates from each of the 500 subsamples provided a "sample" of means and standard deviations from which to estimate national norms more accurately and precisely. At each year level small differences were found between the original means and standard deviations of the achieved sample and the means and standard deviations realised through repeated subsampling. These small differences are reflected in the final published norms.

Table 6 shows sample statistics for the listening tests, by year level. The means and standard deviations reported result from the pos-tstratification exercise. As can be seen, the mean scale score increases fairly uniformly from one year level to the next. This variation is also illustrated in Figure 7 on page 33, which shows the whole distribution of student achievement by year level.

Table 6 Summary statistics for the PAT:Listening Comprehension sample

	Whole sample			Girls			Boys		
Year level	Number of students	Mean (patl)	SD (patl)	Number of students	Mean (patl)	SD (patl)	Number of students	Mean (patl)	SD (patl)
3	1,660	47.3	10.6	835	47.2	10.5	813	47.4	10.6
4	1,623	50.3	11.5	836	50.2	11.5	780	50.4	11.5
5	1,842	52.1	11.0	894	52.1	11.2	945	52.2	10.8
6	1,766	54.4	11.4	883	54.5	11.5	872	54.2	11.3
7	1,679	56.1	11.4	716	55.7	11.4	943	56.5	11.5
8	1,749	58.5	11.4	725	58.6	11.4	1,016	58.4	11.5
9	1,119	63.3	11.7	479	63.1	11.3	634	63.6	12.1
10	1,031	65.4	11.7	470	64.9	11.3	549	65.8	12.1

9.2 Normative information

Stanines

The concept of stanines was introduced in Part A. Here we look more closely at what stanine scores are and how they are calculated.

The word "stanine" is an abridged form of "standard nine". Stanines have been calculated by segmenting the normalised reference distribution of scale scores at each year level into nine equal intervals. At any particular year level, the width of each stanine on the scale (apart from stanines 1 and 9) is equivalent to half the standard deviation of the reference sample for that year level.

It is important to note that although the width of each stanine on the scale is equal, the nine stanines do not hold equal proportions of students. In each year level most students score in stanines 4, 5 and 6, with smaller proportions scoring in stanines 7–9 and stanines 1–3. This can be seen in Figure 8, which shows the relationship of the stanine categories to the normal distribution of scale scores for a particular year level.

Stanine scores for PAT:Listening Comprehension have been calculated for each of Years 3 to 10 using achievement data from the reference sample. Stanines allow students' achievement on PAT:Listening Comprehension to be compared with the achievement of students who made up the national reference samples.

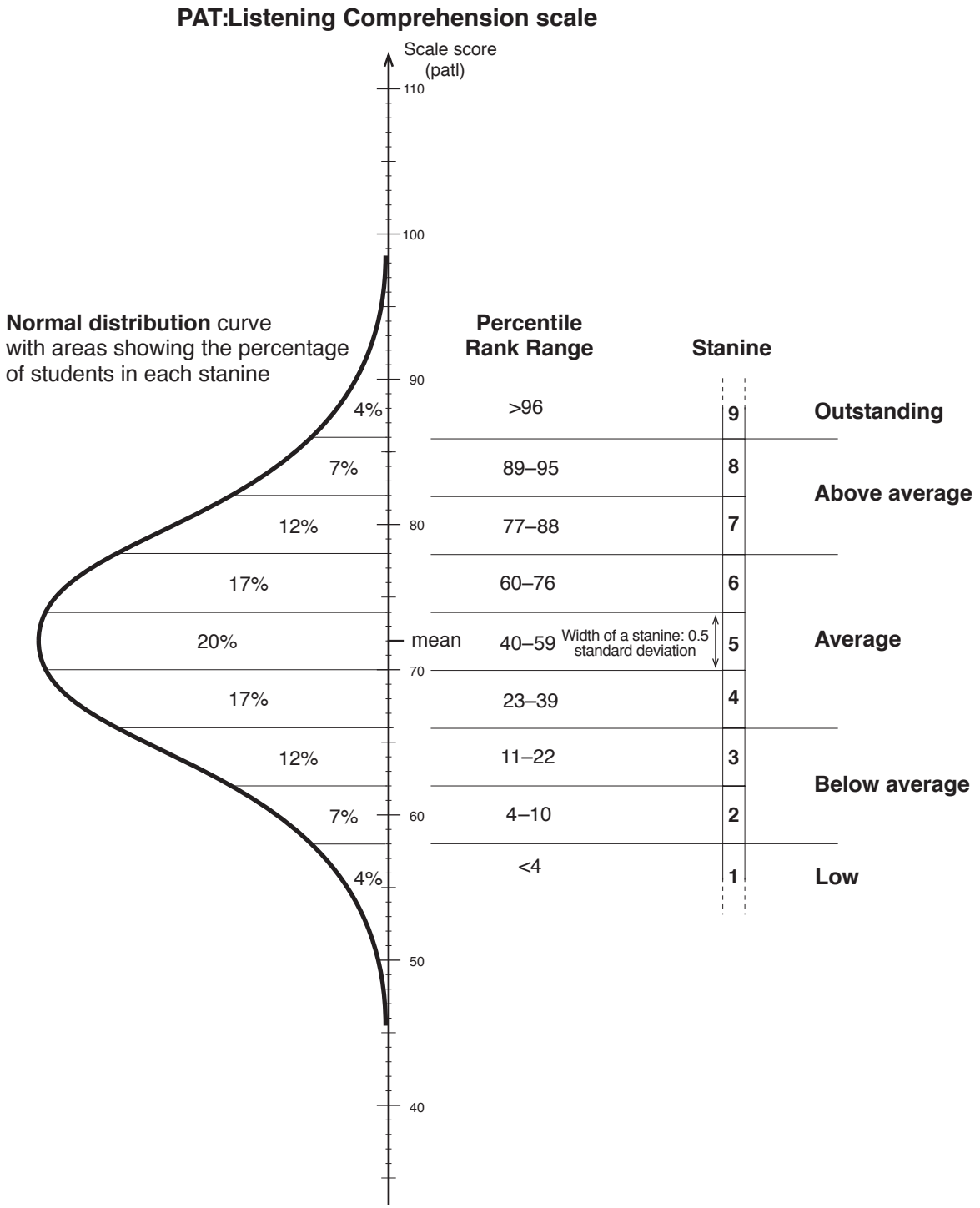
Calculating the stanine scores

Two assumptions have been made in the calculation of stanine scores.

1. The scale scores of the population in each year level are normally distributed.
2. The best estimates of the mean and standard deviation of the normal distribution of scale scores of the population in a year level are the mean and standard deviation, estimated by the post-stratification process described above. The exceptions are for Years 8 and 9, which have been explained above.

In support of these assumptions, the observed distributions of scale scores at each year level were found to be very close to normal distributions. Second, repeated subsampling (as described above), using appropriate post-stratification proportions, was a robust and rigorous process resulting in very stable estimates. Stanine scores have therefore been prepared using theoretical normal distributions based on the modified means and standard deviations for each year level.

Figure 8 Example of stanines and the normal distribution of scale scores on the PATL scale



Interpreting stanines for individual students

Although stanines have valuable properties for interpretation (the stanine is an interval variable), the stanine scale is relatively coarse. When differences in stanines are interpreted between individual students, their scores must differ by at least two stanines before we can be confident that we are observing a true difference in achievement.

Therefore, stanines should be used:

- only as a **broad** measure of achievement.
- when comparing two scales that are not directly comparable (e.g., a student's PAT:Listening Comprehension score can be compared to their PAT:Reading Comprehension score through the stanine scores).

The descriptors used to classify the different stanine categories in Figure 8 are appropriate when considering the performance of an individual student. They allow for the presence of measurement error when categorising student performance as "average" or "above average", for instance. However, stanine descriptors should not be used when describing performance for a group. If, for example, a class of students has an average stanine of 4, it make no sense to report that the class is achieving at average levels or above.

10 Reliability and validity

10.1 Test reliability

The reliability of a test is the degree to which it will provide consistent scores in repeated testing. Reliability coefficients are based on statistical calculations. The more reliable, or precise, a test, the smaller the errors associated with estimated scale scores based on responses to that test.

The reliability coefficient calculated for each of the main norming tests is shown in Table 7. An observed score can be decomposed into a "true" score and an error score. The reliability coefficient is the correlation between observed and true scores. Reliability coefficients normally range from 0 (no relationship between observed and true scores) to +1 (perfect relationship; this is never actually attained in psychological/educational measurement).

The reliability coefficients shown in Table 7 for the eight main norming tests, all between 0.85 and 0.90, are considered satisfactory for this kind of test. They indicate that the tests produce reliable estimates of students' achievement in listening. A reliability of 0.85 means that 85% of the variance in scores is due to real differences between student achievement and only 15% is due to error.

Table 7 Reliability estimates for the PAT:Listening Comprehension norming tests

Trial Test	No. of Students	No. of Items	Mean (%)	SD (%)	Quest Reliability
1	1,660	34	19.5	6.7	0.86
2	800	33	18.3	6.5	0.86
3	883	37	21.9	7.4	0.87
4	870	36	23.6	7.1	0.87
5	792	37	22.9	7.4	0.87
6	876	37	23.7	6.8	0.85
7	447	32	21.6	6.4	0.87
8	425	38	21.8	7.4	0.87

10.2 Precision of scale scores

The estimated scale score for an achievement is the most probable location on the scale that corresponds to a score on a particular test. The error associated with each scale score provides a range of probable values around the estimated location on the scale. We can be reasonably confident that the true location is situated within this range. The best estimate of a true location on the scale for a student is the average of repeated measures of his or her achievement. For each test, scale score errors are reported for each possible scale score. An error of 3.6 patl, for instance, indicates that a given scale score for a student is likely to be within plus or minus 3.6 patl units of the student's "true score" in about 70% of cases.

It is important to note that the error associated with a scale score increases dramatically for students who do extremely well or very poorly on a particular test. This is because in these cases the test provides less information about what these particular students can actually achieve than it would for students who obtained 50% of the maximum test score. A test better targeted to this group's achievement level would provide a more precise indication of their scale score. We recommend that teachers take the scale score error into account when interpreting scores for individual students. A scale score should be understood as an *interval* on the scale instead of a precise point. When two students' scale scores are close together, consider the error of measurement before making any comparisons.

10.3 Validity

The validity of a test is the degree to which the test measures what it was intended to measure. No statistical process can adequately establish whether the construct (in particular, its hierarchical structure and content) is representative of the intended domain. By far the best approach is for a teacher to work systematically through the items of the test. Having done this, he or she is in a sound position to evaluate its suitability for a particular class and assessment objectives.

The PAT:Listening Comprehension scale has been planned and constructed so that it assesses the skills and abilities important to listening comprehension. Each of the test items has been subjected to thorough scrutiny by practising teachers and comprehension experts, and examined by others with expertise in test construction.

In addition, the items included in the tests have all been deemed to fit the Rasch model satisfactorily. This means that the set of items that makes up the eight listening tests can be regarded as indicating achievement on a single underlying construct, which can be qualitatively described on a developmental continuum and measured on an interval scale.

Achievement in listening comprehension improves with the number of years of instruction and/or the extent of exposure to reading material and the world of words. Evidence for validity is therefore provided by the regular increase in achievement on the PAT:Listening Comprehension scale from one year level to the next, as indicated by the distribution of achievement locations on the scale by year level, and the location of the items in the tests that have been administered at each year level (as shown in Figure 7 on page 33.

10.4 Item bias

During the revision of PAT:Listening Comprehension, careful attention was given to gender and ethnic bias. In the first instance, when constructing the tests care was taken to:

- construct tests that contain texts read by a variety of male and female voices
- construct tests that contain texts read by a variety of voice types
- choose texts that reflect New Zealand's culturally diverse population—in each test a variety of cultures is represented
- construct tests that have a balanced gender appeal.

Although texts can offer varying amounts of interest (on average) to different genders, different cultures and different ages, individual items can still be biased regardless of the direction of appeal of the text.

Item bias can be detected in two ways: *facial bias detection* and *statistical bias detection*. The first (facial bias detection) involves an exercise whereby the test development team and the external teacher panel review each item to evaluate the extent of possible bias in the language used. This is essentially an exercise in judgement, and the process is extended to include an evaluation of the texts for the same purpose. For PAT:Listening Comprehension, any suspect items were either modified or excluded.

The second approach is statistical; it exposes or detects differential item functioning (DIF) for two groups of students. In this process, individual items are examined to see whether performance on any of the items is different for any particular subgroup within the national reference sample. DIF analysis was carried out according to both ethnic group and gender. These analyses identified some items that appeared to function differently for different genders or ethnic groups. However, differences between subgroups at the item level were not marked, and there was no noticeable effect overall on the test performance. Items that displayed different difficulty levels for different subgroups were balanced with items showing the opposite effect, so that the final tests were shown to contain no overall bias. As a result, no items were deleted on the basis of DIF.

11 Properties of the published tests

Eight listening tests have been published in this edition of PAT:Listening Comprehension—one test for each year level. Summary statistics for each of these tests are shown in Table 8. As can be seen, the mean item location on the PATL scale for each listening test increases steadily between each year level.

Table 8 Summary statistics for the published PAT:Listening Comprehension tests

Test	No. of Items	Mean Item Location (patl)
1	34	43.0
2	34	45.9
3	35	46.3
4	35	48.8
5	35	49.4
6	35	51.9
7	34	57.6
8	35	61.9

Figure 9 (on page 43) uses percentage scores to show the relative difficulty of the tests. Percentage scores are shown, along with their corresponding scale score. Because the number of items in each test differs, percentage test scores provide a useful method of comparison. Figure 9 shows that Test 1 is the easiest test, with a 50% score at 43.0 patl. Test 2 is slightly more difficult, with a 50% score that is nearly 3 patl units higher than Test 1. The tests become progressively more difficult. The most difficult test is Test 8, with a 50% score at 61.9 patl.

The difference in difficulty between the easiest and most difficult tests is quite large. A percentage score of 70% on Test 1, for instance, represents the same amount of achievement (52 patl) as about a 31% score on Test 8. The tests overlap in difficulty (see Figure 4 on page 30), so that a teacher can choose which test is best suited to a particular class. As we have seen in Part A, when selecting a test to use with a particular group of students it is important to consider its content.

11.1 Percentage of correct answers for each item

A table of observed item facilities by year level cannot be provided for each item because no item has been administered at all year levels. However, when the means and standard deviations of the normalised distributions of achievement for each year level on the PAT:Listening Comprehension scale are known, expected item facilities can be calculated using the Rasch model. These expected item facilities are shown in Figure 10 for all items in the eight tests.

For example, to find the expected percentage correct for item 23 in Test 4 for each year level from 3 to 10, go to Figure 10 and draw a horizontal line at the location of this item (52.9 patl units). Then draw vertical lines through the intersections of the horizontal lines with the curves that belong to each year level. The expected percentage correct can be read on the horizontal axis at the bottom of the figure. This shows that about 53% of Year 6 students are expected to answer this item correctly.

Expected (modelled) percentages correct were compared to observed percentages correct within the year levels available and found to be very close, providing yet another analysis of fit that validates the construction of the scale.

Figure 9 The difficulty of the PAT:Listening Comprehension tests

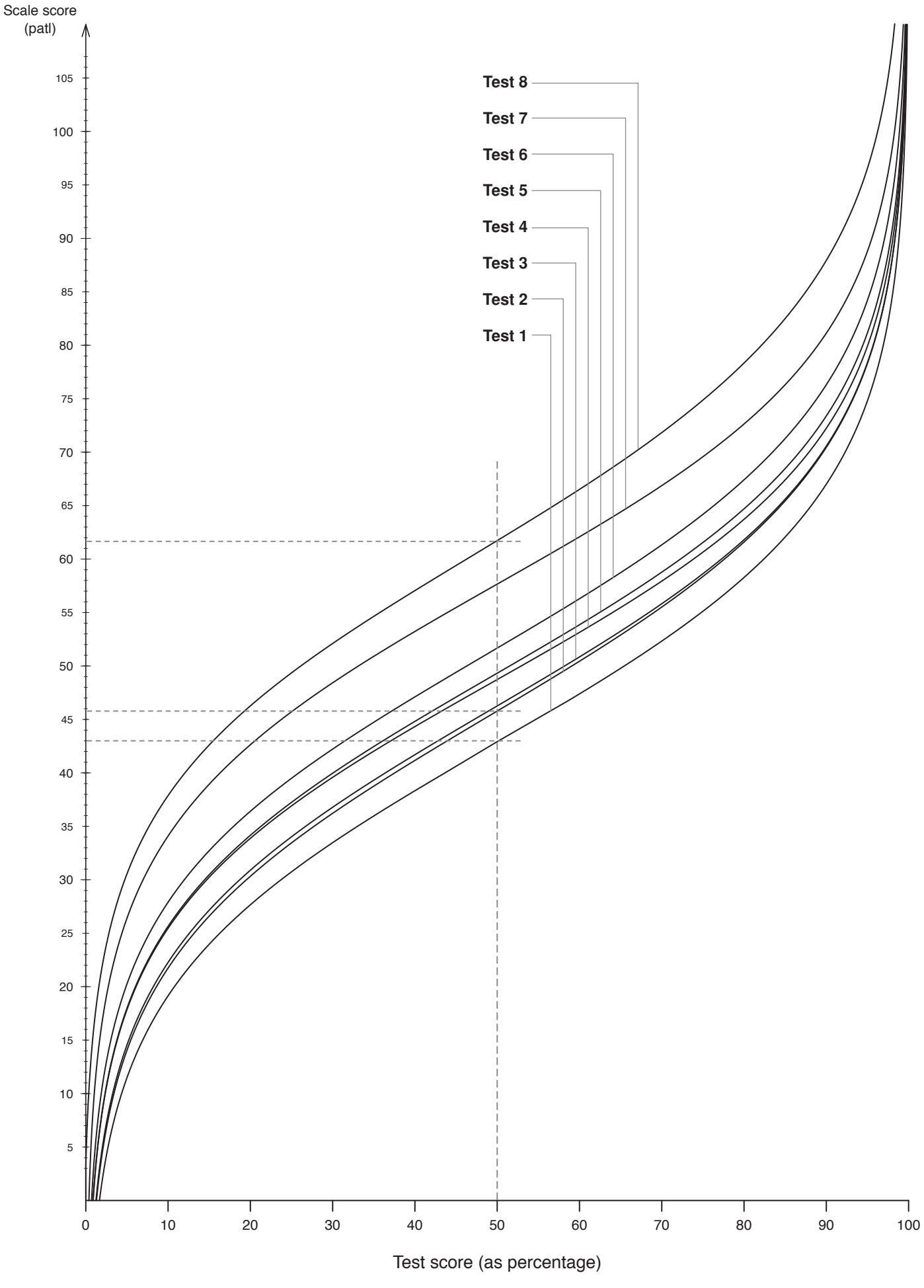
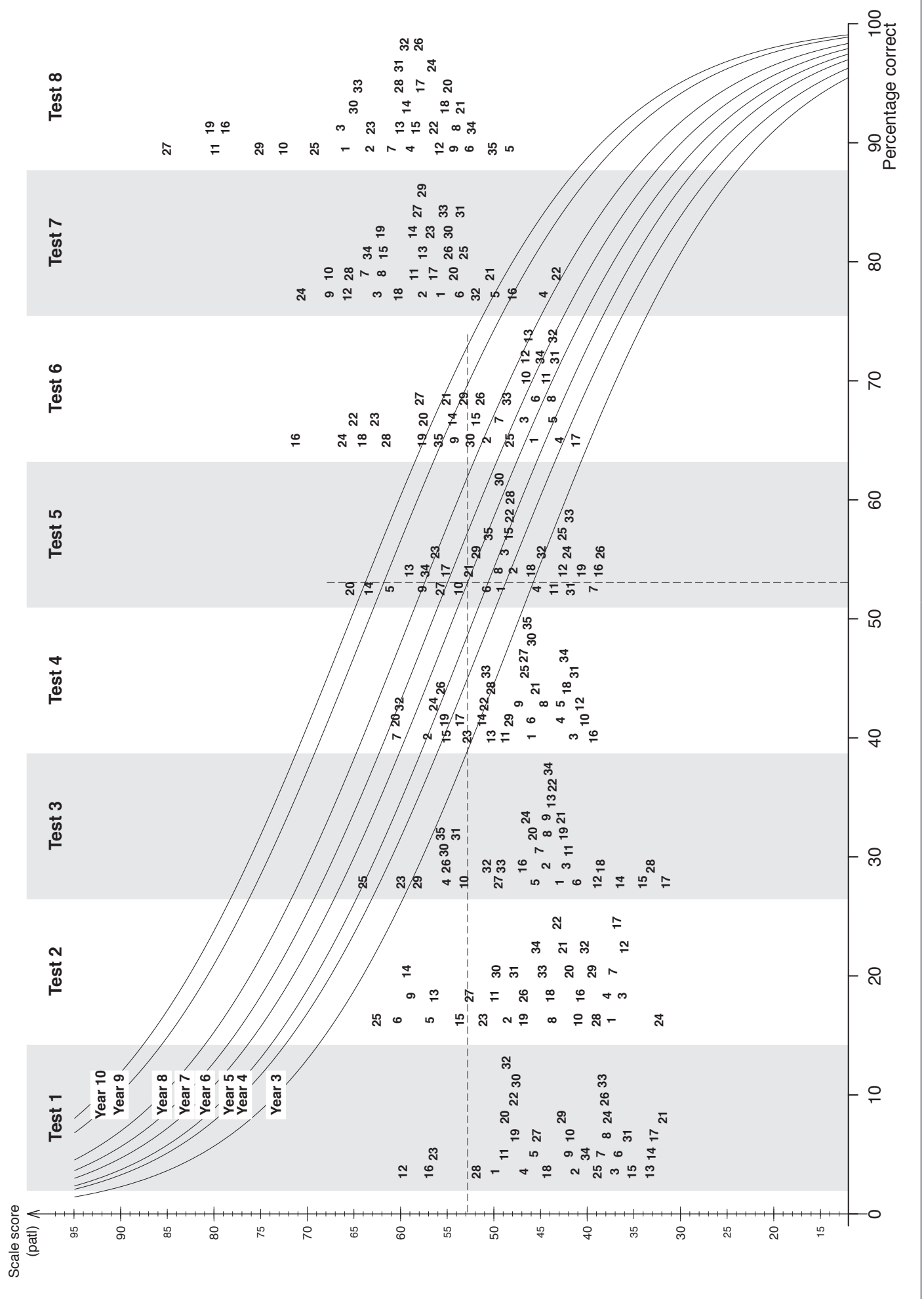


Figure 10 Expected item facility: PAT:Listening Comprehension tests



12 Achievement by age and gender

12.1 Differences in achievement within a year level due to age

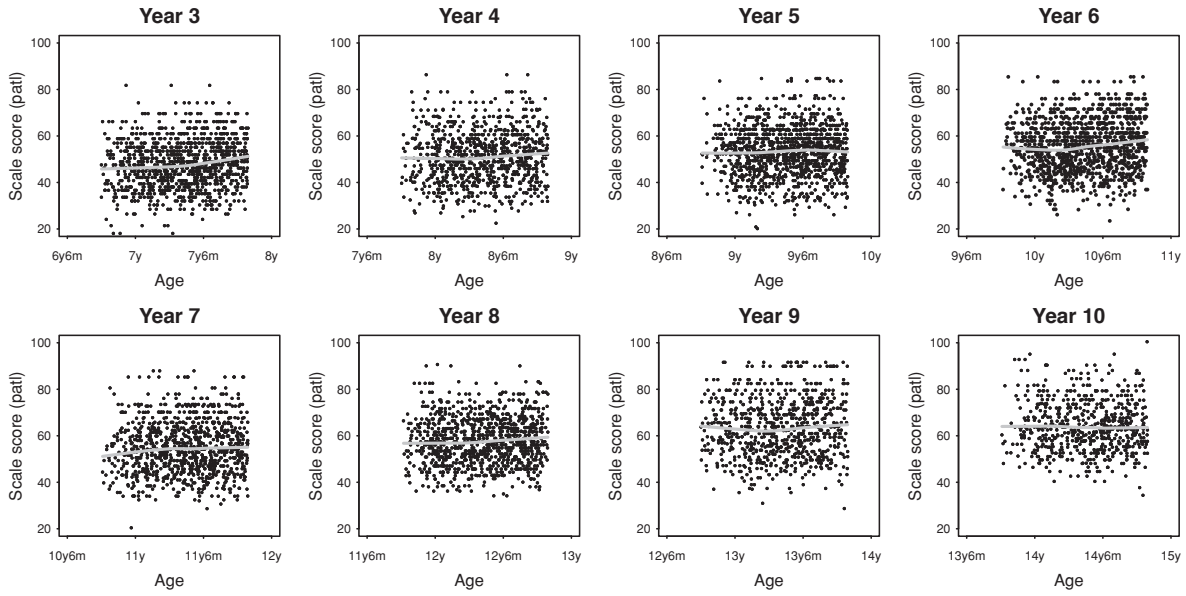
Age-based norms have traditionally been calculated for PAT:Listening Comprehension. The decision not to report these in this edition is based on the following analyses, which show that age effects within separate year groups are not large or consistent enough to warrant norms based on age.

All participating students were asked to provide their date of birth. Correlations between age and scale scores were calculated for each separate year group and across all year levels together. Table 9 shows the Pearson correlation coefficients (r).

The correlation coefficients for the separate year levels are very small, indicating at most a very weak association between age and progress over a school year. Only the correlation for Year 3 shows there may be a lingering "years of schooling" effect at play. The Year 3 correlation of 0.11 means that students with one year more schooling than other students in the same year group may expect, on average, to score 1.1 more patl units on the PAT:Listening Comprehension scale, or approximately one-fifth of a stanine more. However, given the large variability in scores, and taking into consideration measurement error, it would be impossible to say whether a particular score was due to the number of years of schooling, innate ability, or both.

The scatter plots in Figure 11 show the relationship between age and listening scale score by year level.

Figure 11 Correlation of age to scale scores within a year level



Simple linear regressions of scale score against age were run for each year group separately and then for the whole sample. In Table 9 the R^2 value quoted from these models describes the proportion of variance in the scale scores accounted for by age. As can be seen, these percentages are very small indeed within a year group. In other words, age is not a significant factor in scale score variation. The greatest (but minimal) effect is observed in Year 3, where age explains a mere 1% of the variance observed in the scale scores.

Table 9 Correlations between age and scale scores by year level									
	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Year 9	Year 10	Overall
***r* (correlation coefficient)**	0.11	0.07	0.03	0.07	0.06	0.07	0.04	-0.01	0.40
R^2 (proportion of variance in scale scores explained by age)	0.010	0.003	0.000	0.004	0.002	0.004	0.001	0.000	0.160

A general analysis, for the whole sample together, using age to predict scale scores, gives an R^2 of 0.160. A further analysis using year level instead of age to predict scale scores gives a very slightly higher R^2 of 0.167, suggesting that, overall, year level is in fact a slightly better predictor of scale score than age.

12.2 Differences in achievement within year level due to gender

Gender norms are reported in Table 6 on page 37. The differences between average scores by year level for boys and girls are extremely small, and further comparison of distributions according to gender (for instance, stanine graphs or separate conversion tables) is not warranted.

12.3 Correlation between PAT:Listening Comprehension and PAT:Reading Comprehension scores

Schools involved in the PAT:Listening Comprehension trial that had used the NZCER marking service for PAT:Reading Comprehension in Term 1 2010 were asked for permission to compare any results from PAT:Reading Comprehension with the PAT:Listening Comprehension results.

The comparison showed a relationship of moderate strength with comprehension results overall. Correlations within year groups run at approximately 0.6 to 0.8, with a tendency to be on the weaker side in the younger year groups, while showing a slightly stronger relationship in the upper year groups. Overall, comparing all year groups at once the correlation is 0.67.

The comparisons were made with very limited data in some year groups, so these results should be taken as indicators only. Table 10 shows the correlations by year level.

Table 10 Correlation between PAT:Reading Comprehension and PAT:Listening Comprehension scores								
	Year 3	**Year 4**	**Year 5**	**Year 6**	**Year 7**	**Year 8**	**Year 9**	**Year 10**
Correlation	–	0.61	0.67	0.69	0.75	0.70	0.76	0.80

PART C

Conversion Tables and Templates

13 Score conversion tables and associated report templates

Part C contains the various score conversion tables and student reports associated with each of the individual PAT:Listening Comprehension tests.

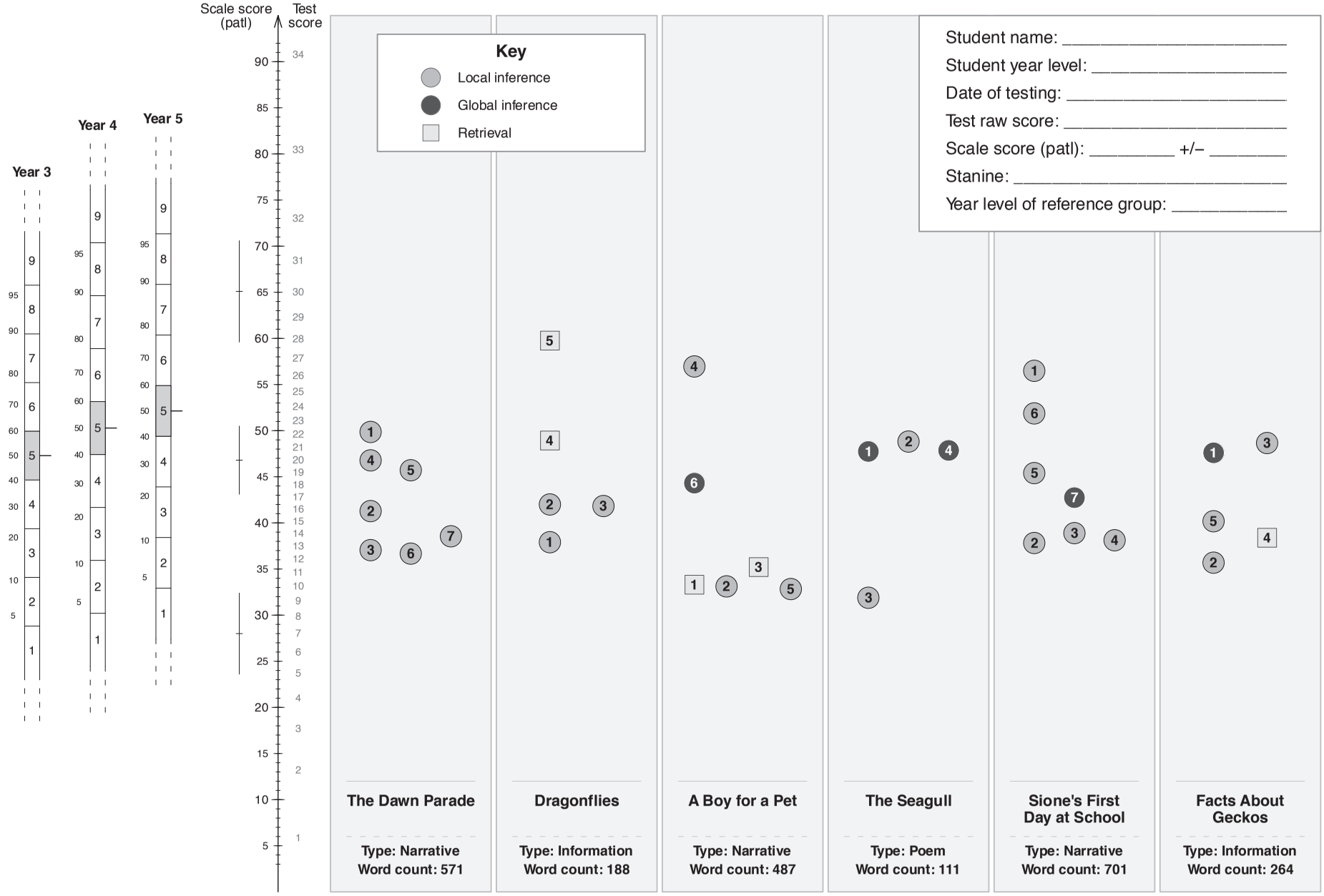
Part A of this manual provides an explanation of the different tables and reports and how to use them. We recommend that teachers and school administrators read Part A before using any of the following tables or reports. Schools are free to copy the student and scale descriptor reports as needed.

LISTENING TEST 1

Score Conversion Table: **Test 1**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 3 stanine	Year 4 stanine	Year 5 stanine
34	90.8	>10.3	9	9	9
33	80.5	10.3	9	9	9
32	73.0	7.4	9	9	9
31	68.5	6.2	9	8	8
30	65.1	5.5	8	8	7
29	62.3	5.0	8	7	7
28	60.0	4.7	7	7	6
27	57.9	4.4	7	6	6
26	56.0	4.2	7	6	6
25	54.3	4.1	6	6	5
24	52.6	4.0	6	5	5
23	51.1	3.9	6	5	5
22	49.6	3.8	5	5	5
21	48.2	3.7	5	5	4
20	46.8	3.7	5	4	4
19	45.5	3.7	5	4	4
18	44.2	3.6	4	4	4
17	42.8	3.6	4	4	3
16	41.5	3.6	4	3	3
15	40.2	3.6	4	3	3
14	38.9	3.7	3	3	3
13	37.5	3.7	3	3	2
12	36.1	3.8	3	3	2
11	34.7	3.8	3	2	2
10	33.2	3.9	2	2	2
9	31.6	4.0	2	2	1
8	29.9	4.2	2	1	1
7	28.0	4.4	1	1	1
6	26.0	4.6	1	1	1
5	23.7	5.0	1	1	1
4	21.0	5.4	1	1	1
3	17.7	6.1	1	1	1
2	13.2	7.4	1	1	1
1	5.9	10.2	1	1	1
0	−4.3	>10.2	1	1	1

PAT: Listening Comprehension Student Report

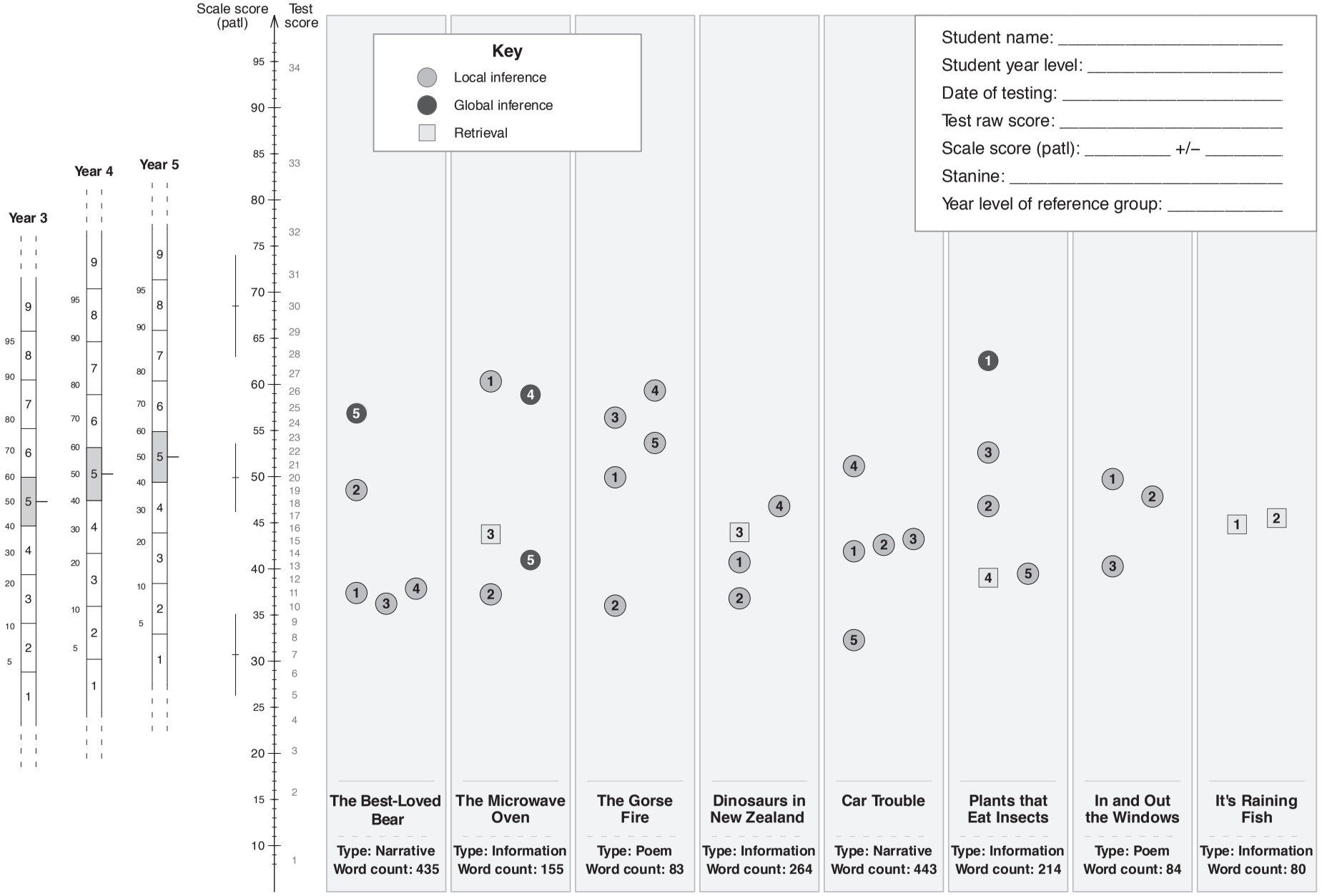


LISTENING TEST 2

Score Conversion Table: **Test 2**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 3 stanine	Year 4 stanine	Year 5 stanine
34	94.3	>10.3	9	9	9
33	84.0	10.3	9	9	9
32	76.5	7.4	9	9	9
31	71.9	6.2	9	9	9
30	68.5	5.5	9	8	8
29	65.7	5.1	8	8	7
28	63.3	4.7	8	7	7
27	61.2	4.5	8	7	7
26	59.3	4.3	7	7	6
25	57.5	4.1	7	6	6
24	55.8	4.0	7	6	6
23	54.3	3.9	6	6	5
22	52.7	3.8	6	5	5
21	51.3	3.8	6	5	5
20	49.9	3.7	5	5	5
19	48.5	3.7	5	5	4
18	47.1	3.7	5	4	4
17	45.8	3.7	5	4	4
16	44.4	3.7	4	4	4
15	43.1	3.7	4	4	3
14	41.7	3.7	4	4	3
13	40.3	3.7	4	3	3
12	38.9	3.8	3	3	3
11	37.4	3.9	3	3	2
10	35.9	4.0	3	2	2
9	34.3	4.1	3	2	2
8	32.6	4.2	2	2	1
7	30.7	4.4	2	2	1
6	28.7	4.7	1	1	1
5	26.4	5.0	1	1	1
4	23.6	5.5	1	1	1
3	20.3	6.2	1	1	1
2	15.8	7.4	1	1	1
1	8.4	10.2	1	1	1
0	−1.8	>10.2	1	1	1

PAT: Listening Comprehension Student Report

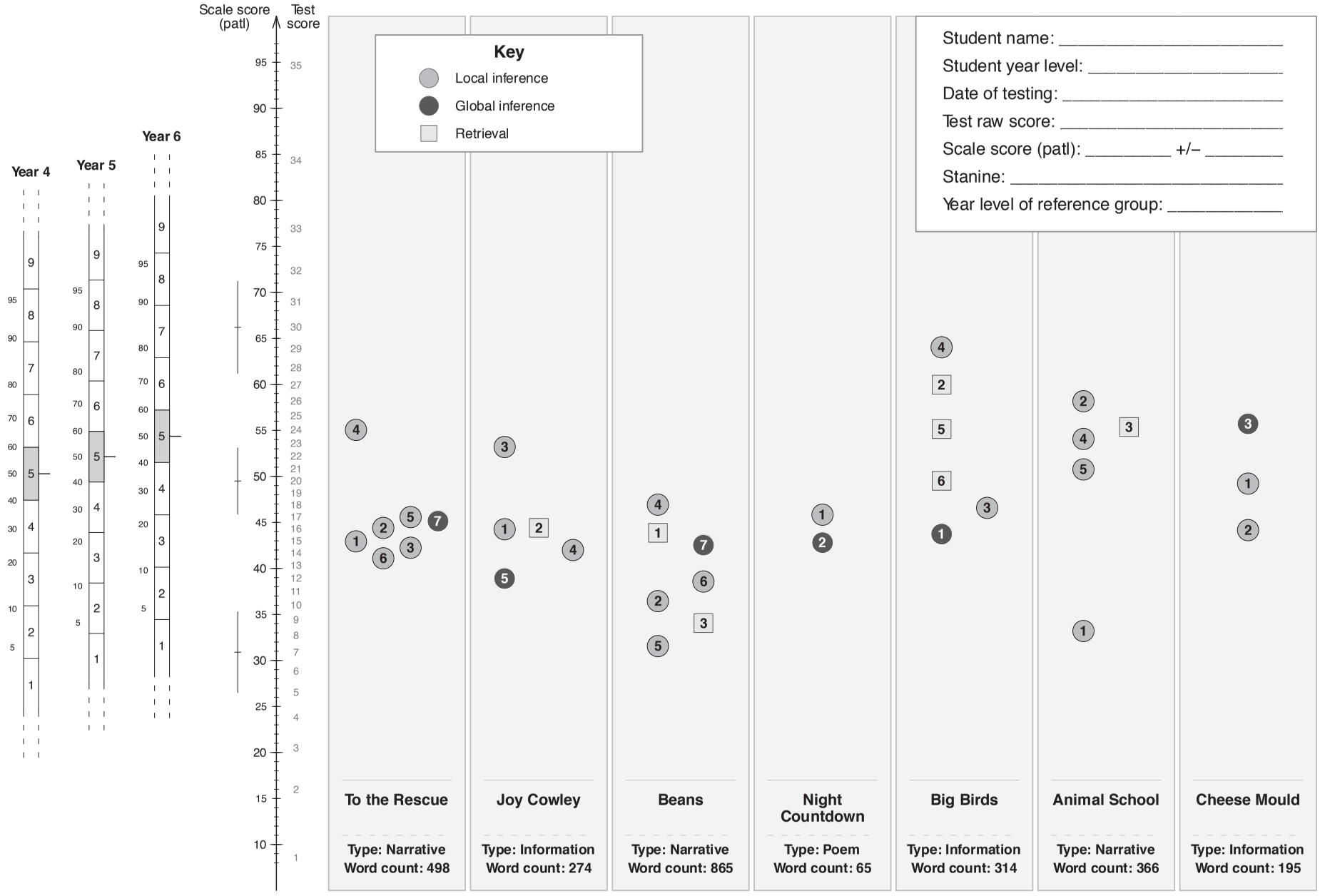


LISTENING TEST 3

Score Conversion Table: **Test 3**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 4 stanine	Year 5 stanine	Year 6 stanine
35	94.7	>10.2	9	9	9
34	84.4	10.2	9	9	9
33	76.9	7.4	9	9	9
32	72.4	6.2	9	9	8
31	69.0	5.5	8	8	8
30	66.2	5.0	8	8	7
29	63.9	4.7	7	7	7
28	61.8	4.4	7	7	6
27	59.9	4.2	7	6	6
26	58.2	4.1	6	6	6
25	56.6	4.0	6	6	5
24	55.1	3.9	6	6	5
23	53.6	3.8	6	5	5
22	52.2	3.7	5	5	5
21	50.8	3.7	5	5	4
20	49.5	3.6	5	5	4
19	48.2	3.6	5	4	4
18	46.9	3.6	4	4	4
17	45.6	3.6	4	4	3
16	44.3	3.6	4	4	3
15	43.0	3.6	4	3	3
14	41.7	3.7	4	3	3
13	40.4	3.7	3	3	3
12	39.0	3.8	3	3	2
11	37.5	3.8	3	2	2
10	36.0	3.9	3	2	2
9	34.4	4.0	2	2	1
8	32.7	4.2	2	1	1
7	30.9	4.4	2	1	1
6	28.9	4.6	1	1	1
5	26.5	5.0	1	1	1
4	23.8	5.5	1	1	1
3	20.5	6.2	1	1	1
2	16.0	7.4	1	1	1
1	8.6	10.2	1	1	1
0	−1.7	>10.2	1	1	1

PAT: Listening Comprehension Student Report

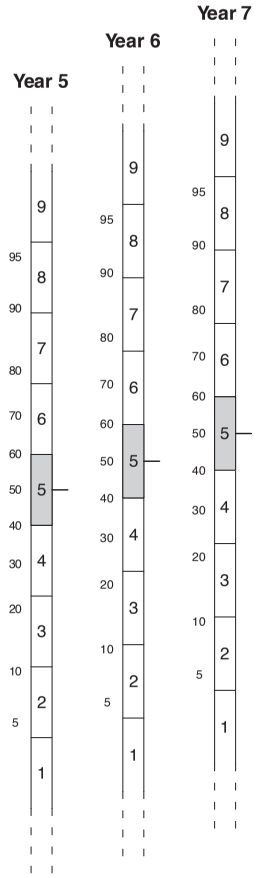


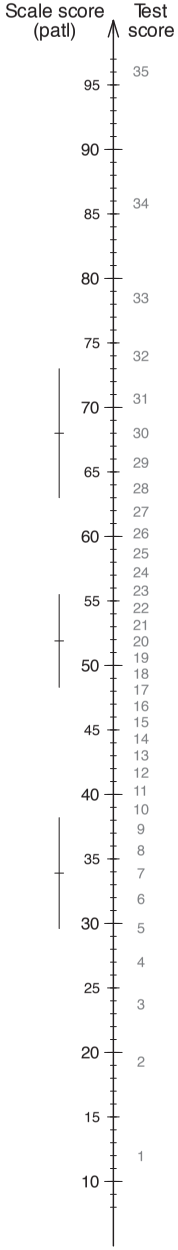
LISTENING TEST 4

Score Conversion Table: **Test 4**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 5 stanine	Year 6 stanine	Year 7 stanine
35	96.1	>10.2	9	9	9
34	85.8	10.2	9	9	9
33	78.5	7.4	9	9	9
32	74.0	6.1	9	8	8
31	70.7	5.4	8	8	7
30	68.0	5.0	8	7	7
29	65.7	4.6	7	7	7
28	63.7	4.4	7	7	6
27	61.9	4.2	7	6	6
26	60.2	4.0	6	6	6
25	58.7	3.9	6	6	5
24	57.2	3.8	6	5	5
23	55.8	3.7	6	5	5
22	54.5	3.6	5	5	5
21	53.1	3.6	5	5	4
20	51.9	3.6	5	5	4
19	50.6	3.5	5	4	4
18	49.4	3.5	5	4	4
17	48.1	3.5	4	4	4
16	46.9	3.5	4	4	3
15	45.6	3.6	4	3	3
14	44.3	3.6	4	3	3
13	43.0	3.6	3	3	3
12	41.7	3.7	3	3	2
11	40.3	3.8	3	3	2
10	38.8	3.9	3	2	2
9	37.3	4.0	2	2	2
8	35.7	4.1	2	2	1
7	33.9	4.3	2	1	1
6	31.9	4.6	1	1	1
5	29.6	4.9	1	1	1
4	27.0	5.4	1	1	1
3	23.7	6.1	1	1	1
2	19.3	7.3	1	1	1
1	12.0	10.2	1	1	1
0	1.8	>10.2	1	1	1

PAT: Listening Comprehension Student Report





Student name: _____

Student year level: _____

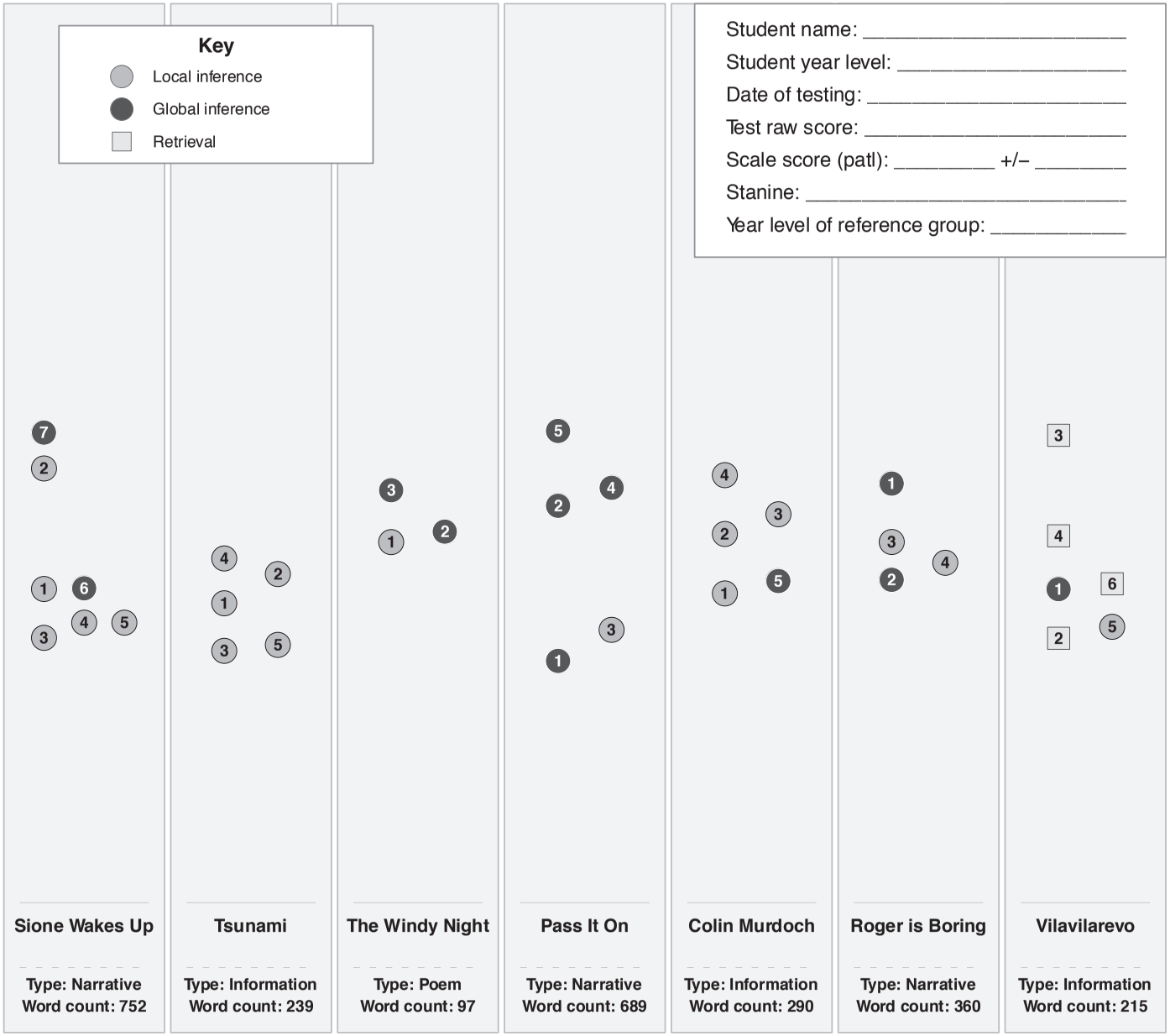
Date of testing: _____

Test raw score: _____

Scale score (patl): _____ +/– _____

Stanine: _____

Year level of reference group: _____



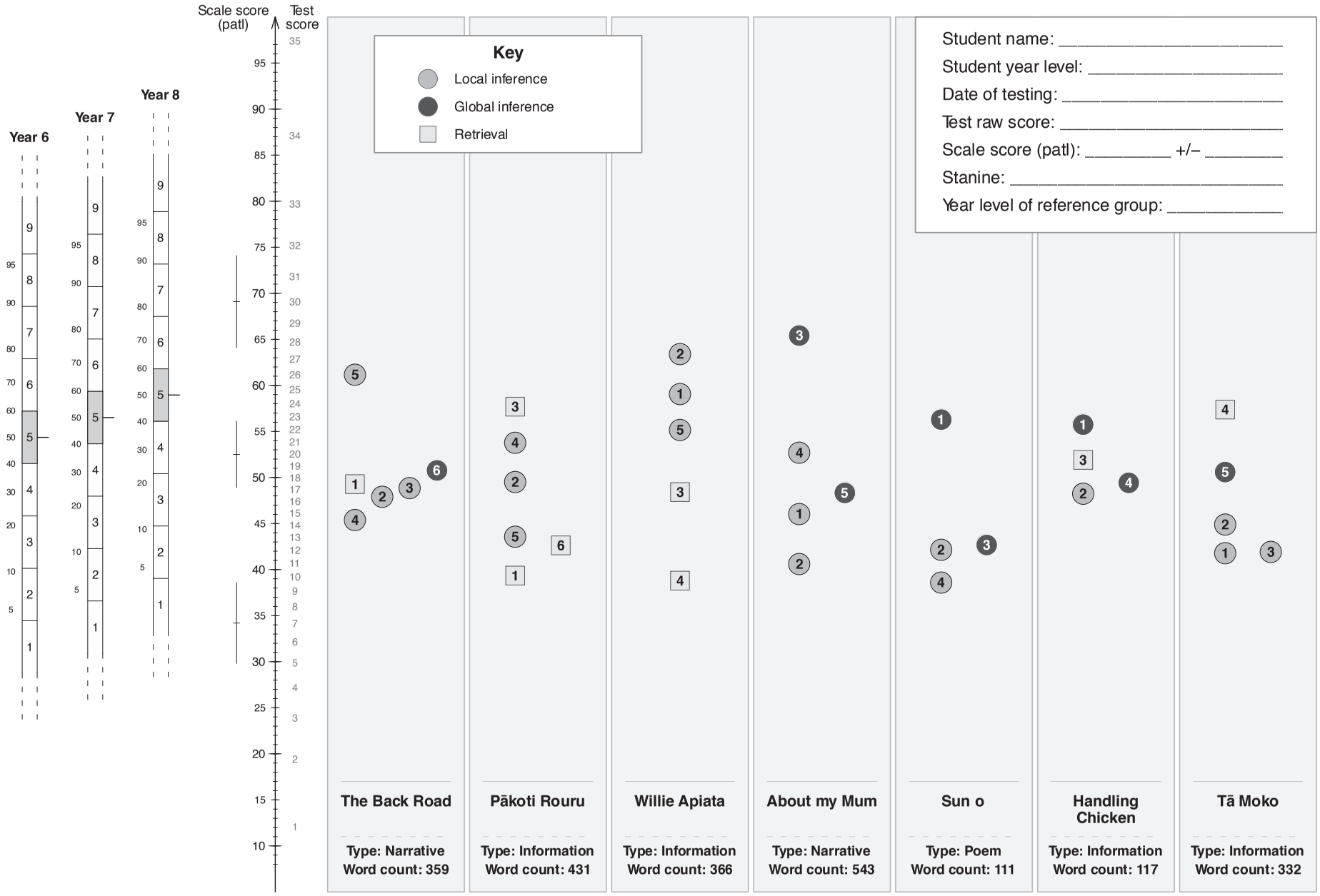
Sione Wakes Up	Tsunami	The Windy Night	Pass It On	Colin Murdoch	Roger is Boring	Vilavilarevo
Type: Narrative Word count: 752	Type: Information Word count: 239	Type: Poem Word count: 97	Type: Narrative Word count: 689	Type: Information Word count: 290	Type: Narrative Word count: 360	Type: Information Word count: 215

LISTENING TEST 5

Score Conversion Table: **Test 5**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 6 stanine	Year 7 stanine	Year 8 stanine
35	97.4	>10.2	9	9	9
34	87.1	10.2	9	9	9
33	79.7	7.4	9	9	9
32	75.2	6.2	9	8	8
31	71.8	5.5	8	8	7
30	69.1	5.0	8	7	7
29	66.8	4.7	7	7	6
28	64.7	4.4	7	6	6
27	62.8	4.2	6	6	6
26	61.1	4.1	6	6	5
25	59.5	3.9	6	6	5
24	58.0	3.8	6	5	5
23	56.6	3.8	5	5	5
22	55.2	3.7	5	5	4
21	53.9	3.6	5	5	4
20	52.5	3.6	5	4	4
19	51.2	3.6	4	4	4
18	50.0	3.6	4	4	3
17	48.7	3.6	4	4	3
16	47.4	3.6	4	3	3
15	46.1	3.6	4	3	3
14	44.8	3.6	3	3	3
13	43.5	3.7	3	3	2
12	42.1	3.7	3	2	2
11	40.7	3.8	3	2	2
10	39.2	3.9	2	2	2
9	37.6	4.0	2	2	1
8	36.0	4.2	2	1	1
7	34.2	4.4	1	1	1
6	32.1	4.6	1	1	1
5	29.9	4.9	1	1	1
4	27.2	5.4	1	1	1
3	23.9	6.1	1	1	1
2	19.4	7.4	1	1	1
1	12.1	10.2	1	1	1
0	1.9	>10.2	1	1	1

PAT: Listening Comprehension Student Report

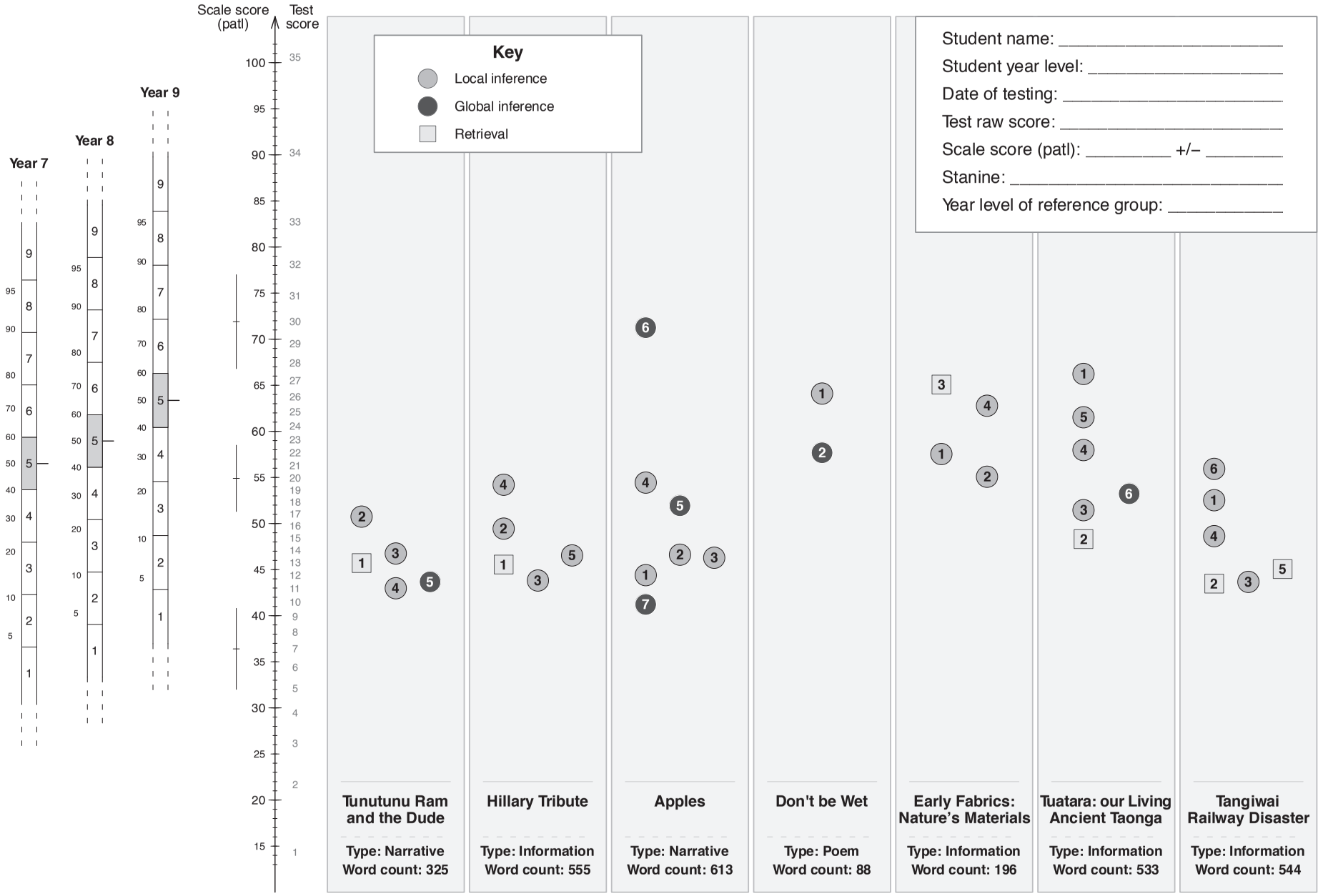


LISTENING TEST 6

Score Conversion Table: **Test 6**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 7 stanine	Year 8 stanine	Year 9 stanine
35	100.6	>10.3	9	9	9
34	90.2	10.3	9	9	9
33	82.7	7.5	9	9	8
32	78.1	6.2	9	8	8
31	74.7	5.5	8	8	7
30	71.9	5.1	8	7	6
29	69.5	4.7	7	7	6
28	67.4	4.5	7	6	6
27	65.5	4.3	7	6	5
26	63.8	4.1	6	6	5
25	62.1	4.0	6	6	5
24	60.6	3.9	6	5	5
23	59.1	3.8	5	5	4
22	57.7	3.7	5	5	4
21	56.3	3.7	5	5	4
20	54.9	3.6	5	4	4
19	53.6	3.6	4	4	3
18	52.3	3.6	4	4	3
17	51.0	3.6	4	4	3
16	49.7	3.6	4	3	3
15	48.4	3.6	4	3	2
14	47.1	3.6	3	3	2
13	45.8	3.7	3	3	2
12	44.4	3.7	3	2	2
11	43.0	3.8	3	2	2
10	41.5	3.9	2	2	1
9	39.9	4.0	2	2	1
8	38.2	4.2	2	1	1
7	36.4	4.4	1	1	1
6	34.4	4.6	1	1	1
5	32.1	4.9	1	1	1
4	29.5	5.4	1	1	1
3	26.1	6.1	1	1	1
2	21.7	7.4	1	1	1
1	14.3	10.2	1	1	1
0	4.1	>10.2	1	1	1

PAT: Listening Comprehension Student Report

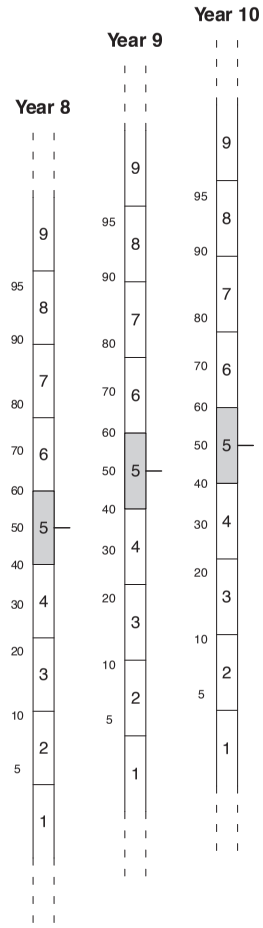


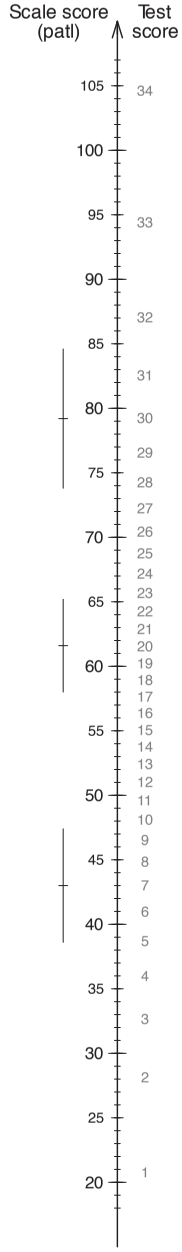
LISTENING TEST 7

Score Conversion Table: **Test 7**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 8 stanine	Year 9 stanine	Year 10 stanine
34	104.6	>10.2	9	9	9
33	94.4	10.2	9	9	9
32	87.0	7.4	9	9	9
31	82.6	6.1	9	8	8
30	79.2	5.4	9	8	7
29	76.6	5.0	8	7	7
28	74.3	4.6	8	7	7
27	72.3	4.4	7	7	6
26	70.4	4.2	7	6	6
25	68.7	4.0	7	6	6
24	67.2	3.9	6	6	5
23	65.7	3.8	6	5	5
22	64.3	3.7	6	5	5
21	62.9	3.7	6	5	5
20	61.6	3.6	5	5	4
19	60.2	3.6	5	4	4
18	58.9	3.6	5	4	4
17	57.7	3.6	5	4	4
16	56.4	3.6	5	4	3
15	55.1	3.6	4	4	3
14	53.8	3.6	4	3	3
13	52.4	3.7	4	3	3
12	51.0	3.7	4	3	3
11	49.6	3.8	3	3	2
10	48.1	3.9	3	2	2
9	46.5	4.0	3	2	2
8	44.9	4.2	3	2	2
7	43.0	4.4	2	2	1
6	41.0	4.6	2	1	1
5	38.7	5.0	1	1	1
4	36.0	5.4	1	1	1
3	32.7	6.2	1	1	1
2	28.2	7.4	1	1	1
1	20.8	10.2	1	1	1
0	10.5	>10.2	1	1	1

PAT: Listening Comprehension Student Report





Student name: _____

Student year level: _____

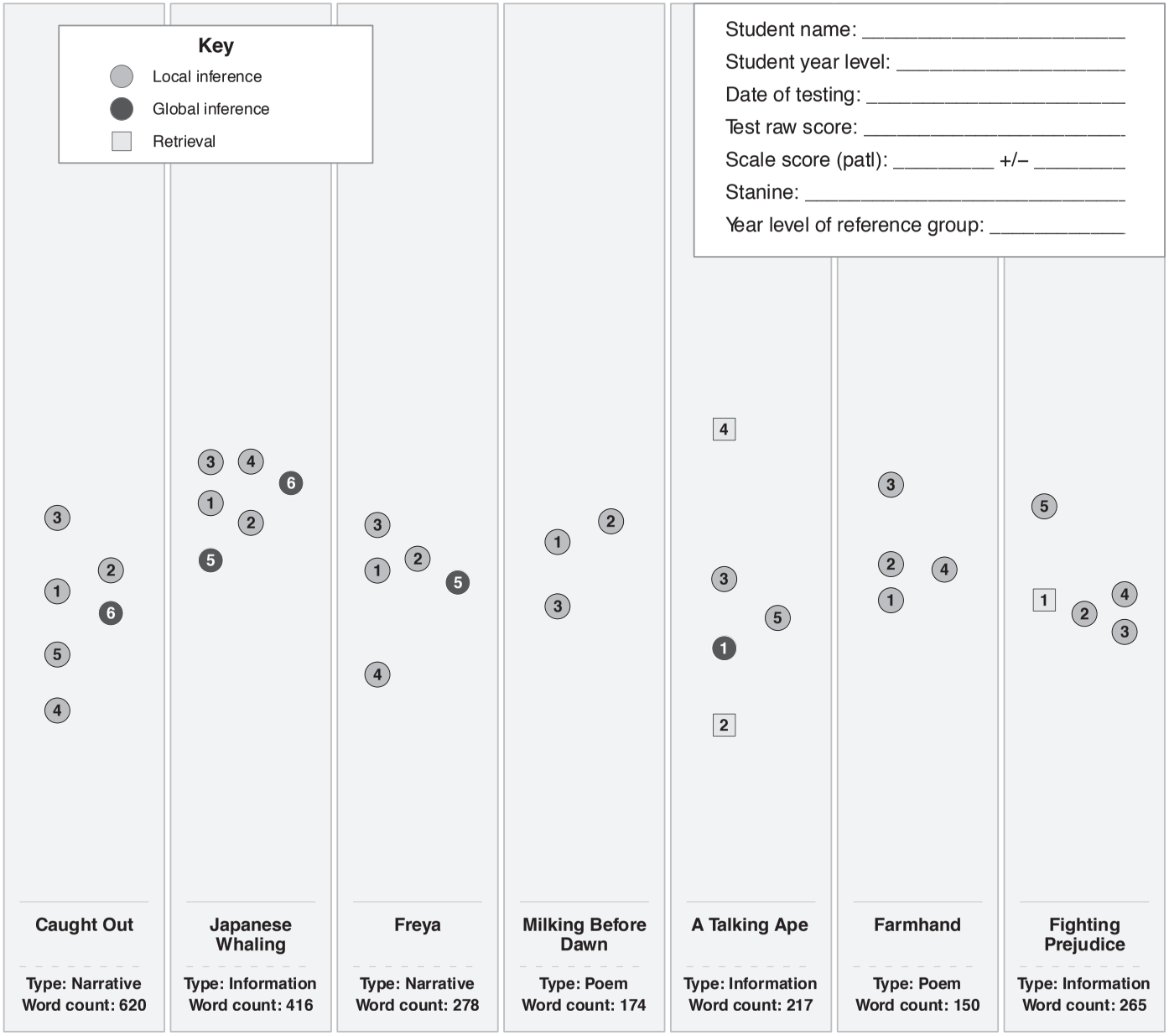
Date of testing: _____

Test raw score: _____

Scale score (patl): _____ +/– _____

Stanine: _____

Year level of reference group: _____



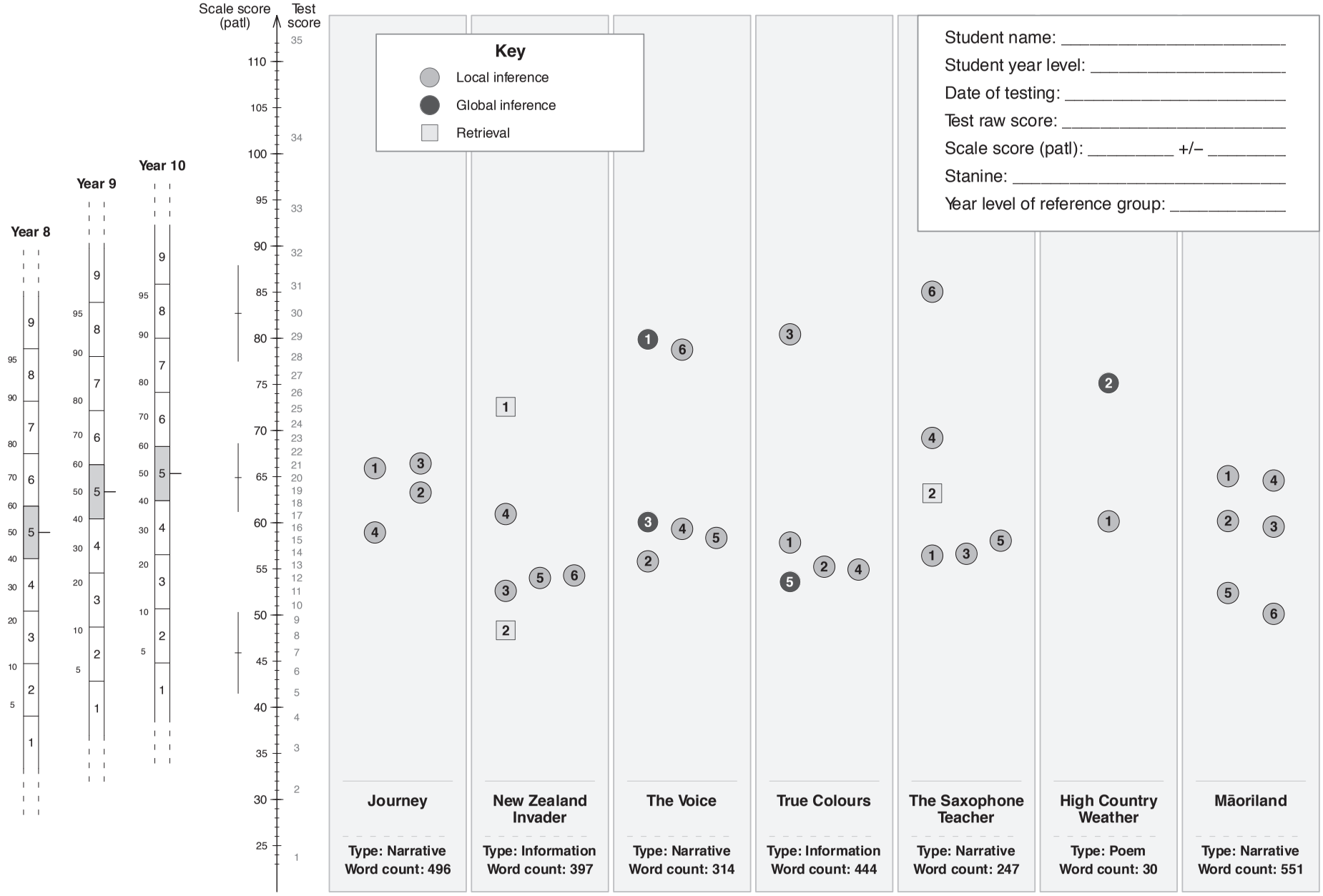
Caught Out	Japanese Whaling	Freya	Milking Before Dawn	A Talking Ape	Farmhand	Fighting Prejudice
Type: Narrative	Type: Information	Type: Narrative	Type: Poem	Type: Information	Type: Poem	Type: Information
Word count: 620	Word count: 416	Word count: 278	Word count: 174	Word count: 217	Word count: 150	Word count: 265

LISTENING TEST 8

Score Conversion Table: **Test 8**

Test score (number correct)	Scale score (patl)	Error (patl)	Year 8 stanine	Year 9 stanine	Year 10 stanine
35	112.3	>10.4	9	9	9
34	101.8	10.4	9	9	9
33	94.1	7.6	9	9	9
32	89.3	6.4	9	9	9
31	85.7	5.7	9	9	8
30	82.7	5.2	9	8	8
29	80.2	4.9	9	8	8
28	78.0	4.6	8	7	7
27	76.0	4.4	8	7	7
26	74.1	4.2	8	7	6
25	72.4	4.1	7	7	6
24	70.7	4.0	7	6	6
23	69.2	3.9	7	6	6
22	67.7	3.8	7	6	5
21	66.3	3.8	6	6	5
20	64.9	3.7	6	5	5
19	63.5	3.7	6	5	5
18	62.2	3.7	6	5	4
17	60.8	3.6	5	5	4
16	59.5	3.6	5	4	4
15	58.2	3.7	5	4	4
14	56.8	3.7	5	4	4
13	55.4	3.7	4	4	3
12	54.0	3.8	4	3	3
11	52.6	3.8	4	3	3
10	51.1	3.9	4	3	3
9	49.5	4.0	3	3	2
8	47.8	4.2	3	2	2
7	45.9	4.4	3	2	2
6	43.9	4.6	2	2	1
5	41.6	5.0	2	1	1
4	38.9	5.4	1	1	1
3	35.6	6.1	1	1	1
2	31.1	7.4	1	1	1
1	23.8	10.2	1	1	1
0	13.5	>10.2	1	1	1

PAT: Listening Comprehension Student Report



14 PATL scale descriptions

The PAT:Listening Comprehension scale is described in seven different ways:

- using abstract information
- using separated information
- using multiple pieces of information
- using implied information
- rejecting competing information
- using lexical information (vocabulary)
- using prosodic information.[5]

This is not an exhaustive list. However, it covers elements that have a significant role to play in listening comprehension.

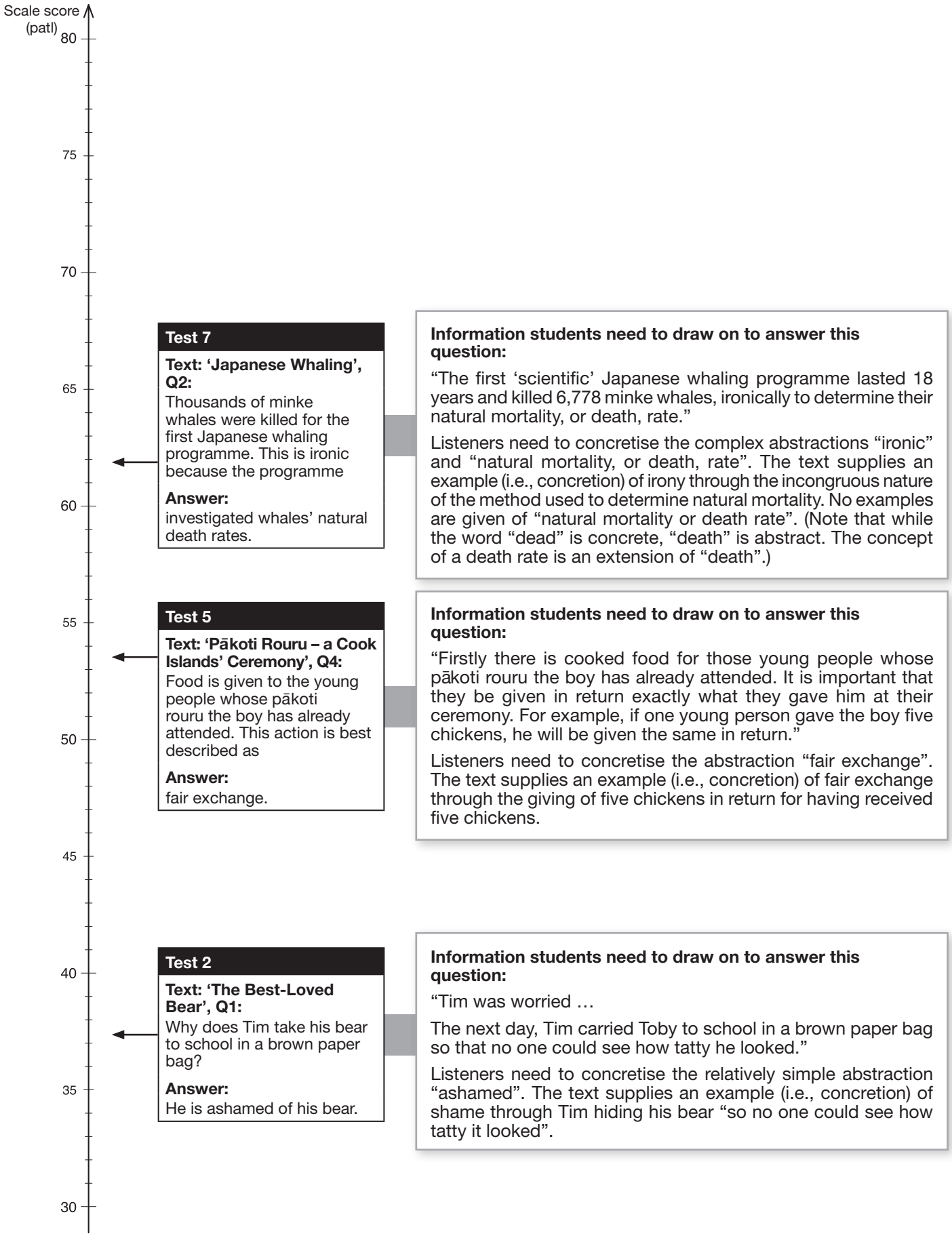
For each element, example questions are located on the scale to exemplify increasing complexity. The text used to correctly answer each question is quoted, or referred to more generally if it is extensive. An analysis of the relationship between the question and text is then presented in order to reveal the kinds of thinking the students are required to do.

The analysis presented in the scale descriptions will provide teachers with a model of how to analyse their own student data so that student strengths and weaknesses can be more deeply understood. For example, teachers might use the descriptions to learn more about each of the seven elements and then use what they learn to better understand why particular students had difficulty with particular questions. This understanding would then be used to identify any patterns of difficulty across groups of students.

5 Prosodic questions assess the student's ability to appreciate prosodic features—the patterns of stress and intonation in a language (stress is the relative emphasis given to certain syllables in a word; intonation is the variation in voice pitch). These questions require students to pay attention to the sound of the narrator's voice in order to comprehend emotions such as enthusiasm, frustration, delight, certainty and excitement. Note that it is possible to *hear* when someone's teeth are nearly clenched or their lips are pursed when they speak; seeing the speaker is not necessary. Note also that in written language we use punctuation to capture some of the stress and intonation of spoken language.

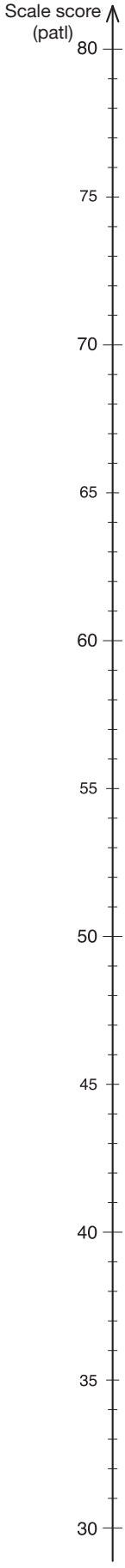
Using abstract information

The examples below range from more complex abstract information (62.0 patl) to less complex abstract information (37.4 patl).



Using separated information

The examples below range from more complex separated information (80.0 patl) to less complex separated information (47.8 patl).



Test 8

Text: 'The Voice', Q1:
The rodent voice inside Frankie's head is best described as the voice of

Answer:
a worrier.

Information students need to draw on to answer this question:

"the rodent voice inside his head told him this was probably a brain aneurism, just as minutes ago it had insisted his aching muscles were meningitis … He had been mentally reciting batting shots … to shut out the future, which seemed to be worsening by the day … he wanted to cease worrying ".

Listeners need to synthesise incomplete information about the rodent voice from across the whole text. The rodent voice, which is never explicitly explained, is mentioned in the second, third and fourth paragraphs, becoming progressively clearer as information accrues. A statement near the end summarises Frankie's feelings: "he wanted to cease worrying".

Test 8

Text: 'New Zealand Invader', Q4:
New Zealand mud snails are best described as

Answer:
tough.

Information students need to draw on to answer this question:

"found in just about any sort of freshwater … It can stand a lick of salt … Even where other native organisms are knocked back by dirty runoff, effluent or silt, it usually survives … it can survive out of water for a time—certainly hours, probably days ….

The snail is an unfussy eater …

a hardy … gastropod."

Listeners need to synthesise information separated across three paragraphs. Although listeners hear the word "hardy" near the end, it is likely that this word functions mainly to confirm an inference of a robust animal already established in the listener's mind by that point. The information used to build this inference is directly and unambiguously stated, but made more complex by being spread across a text full of technical vocabulary.

Test 4

Text: 'The Windy Night', Q2:
Who does the first voice most likely belong to?

Answer:
A child.

Information students need to draw on to answer this question:

"Just branches, / child …

Why, Grandma's / moonlight cat / does that …

that's all, my child".

Listeners need to synthesise information from across the whole text, which is conveyed by alternate voices. They must identify two separate voices in conversation and come to a conclusion regarding identity through explicit statements (e.g., "Just branches / child"), and by inferring that the frightened nocturnal questions are typical of a child. The word "child" is likely to act as a scaffold but is unlikely to be sufficient on its own.

Test 1

Text: 'The Seagull', Q1:
The seagull tells us what he sees from

Answer:
high in the sky.

Information students need to draw on to answer this question:

"Seagull, seagull / Riding high …

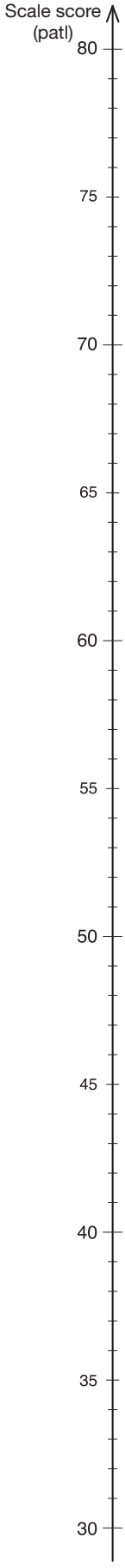
The men and the women / And girls and boys / Look far, far smaller / Than painted toys.

I like to glide / On my wings and stare."

Listeners need to synthesise information from across most of the poem. Many of the single-sentence stanzas that make up the seagull's response confirm its position high in the sky by either describing a view from a high place or by describing the act of flight itself.

Using multiple pieces of information

The examples below range from more complex multiple pieces of information (58.6 patl) to less complex multiple pieces of information (42.63 patl).



Test 7

Text: 'Japanese Whaling', Q5:
What is the main purpose of the text?

Answer:
To criticise scientific whaling

Information students need to draw on to answer this question:

Listeners need to integrate many pieces of information (too many to state here). This information often comes in the form of technical vocabulary (such as "non-lethal methods" and "contaminants"). This vocabulary is used as part of multiclause sentences, e.g., "This makes lethal methods particularly unsuitable for studying whale behaviour, such as migration, which is of great interest to scientists." Here the embedded clause, "such as migration" is used to insert an example of whale behaviour.

Test 6

Text: 'Hillary Tribute', Q2:
How did Hillary prove that he and Tenzing reached the summit?

Answer:
He took photos down the ridges.

Information students need to draw on to answer this question:

"Seconds later Tenzing Norgay joined Hillary on the summit … Hillary took up his camera … he took shots down each of the major ridges as proof of the climb."

Listeners need to integrate several pieces of information. There is some technical vocabulary (such as "summit" and "ridge"), but it will be relatively familiar, and contextual clues are present. This information comes in the form of multiclause sentences (e.g., "Then, peeling off his mitts and outer gloves, Hillary took up his camera in silk-gloved hands and snapped shots of his friend holding an ice-axe strung with flags: Nepalese, Indian, United Nations, the Union Jack—it was, after all, a British expedition").

Test 2

Text: 'Car Trouble', Q2:
Why do Carl's hands sweat?

Answer:
He is worried his Mum will realise what he did.

Information students need to draw on to answer this question:

"Carl doesn't know how he could have forgotten. He spent enough time on it! …

It is when Mum pulls out into the street and changes gear that Carl remembers. There goes the sound: *purrrppptttta!*

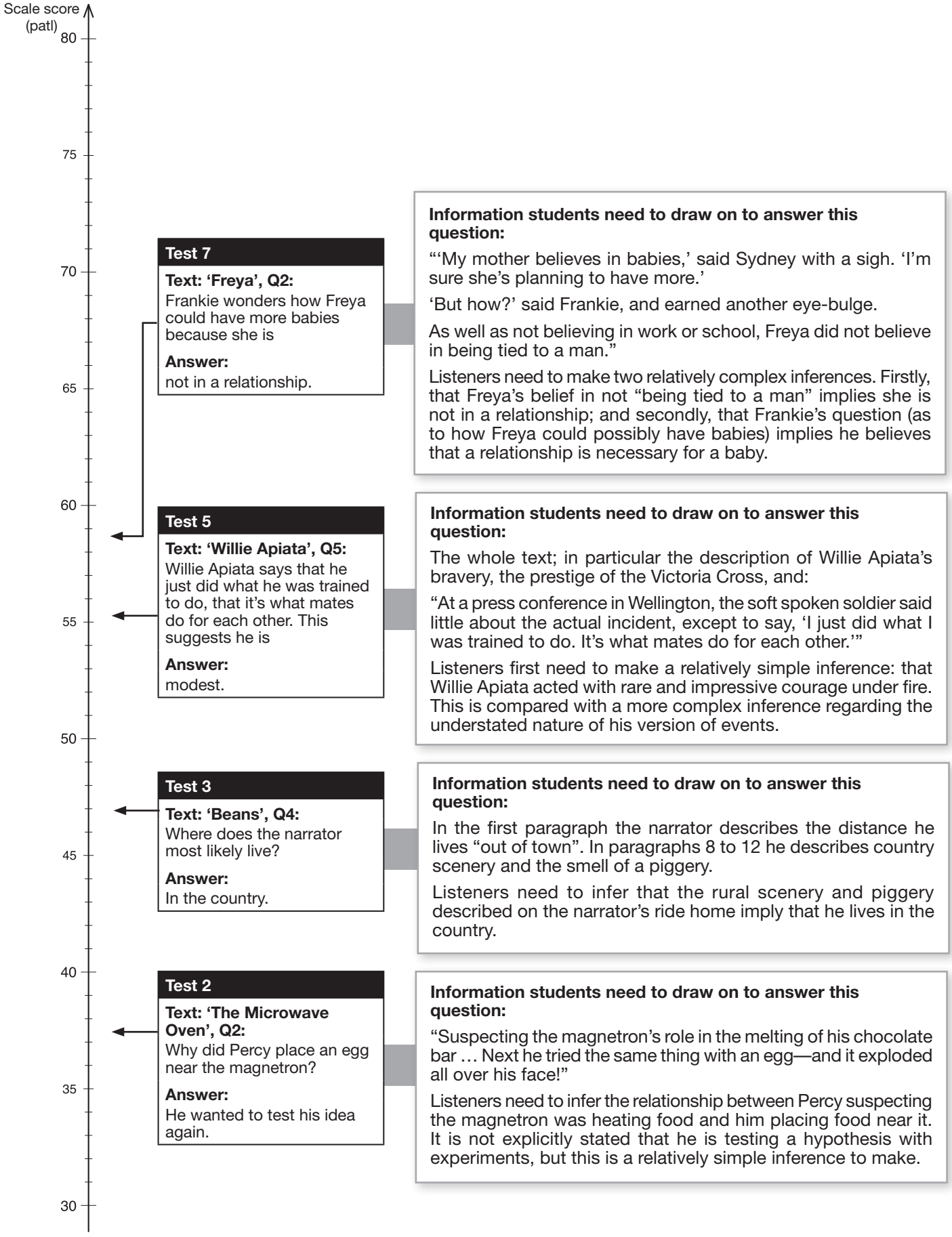
The bicycle tube around the exhaust pipe of the car. *Purrrppptttta … prttt … prt-t-t-t …*

'Can you guys hear something?' Mum asks, frowning into the rear vision mirror. 'Is that us?' 'Don't think so,' Carl says, his hands beginning to sweat. *Please come off*, he prays to the rubber tube. He just knows Mum would *not* find this prank at *all* funny."

Listeners need to integrate several pieces of information. Vocabulary that could be described as technical ("bicycle tube") is present, but a precise definition is not necessary for comprehension. Sentences are uncomplicated.

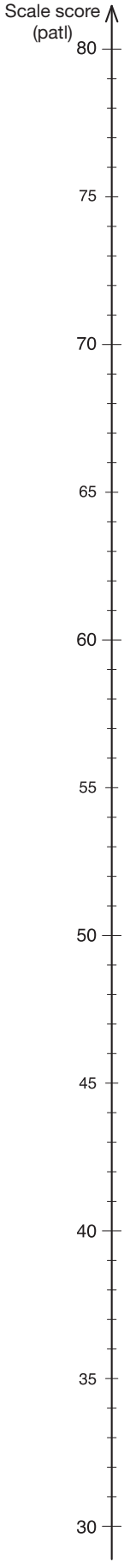
Using implied information

The examples below range from more complex implied information (58.7 patl) to less complex implied information (37.3 patl).



Rejecting competing information

The examples below range from more complex competing information (63.3 patl) to less complex competing information (40.8 patl).



Test 8

Text: 'Journey', Q2:
How does the narrator feel about Māori men using Pākehā machines to cut up the land?

Answer:
Accepting.

Information students need to draw on to answer this question:

"but there you are, a man had to eat, had to get from here to there—anyone knew that."

Competing information that could distract listeners:

"Funny people those Pākehās, had to chop up everything … Couldn't give life, only death."

Listeners need to reject information that might lead them to expect the Māori narrator would be angry or depressed about Māori using Pākehā machines to cut up the land. The competing information offers a very reasonable expectation, which the narrator subverts by quickly moving to pragmatic acceptance.

Test 3

Text: 'Animal School', Q5:
At the end of the story, Lambkins learns to

Answer:
stay in character.

Information students need to draw on to answer this question:

"'Liar!' snarled Lambkins, leaping to his hooves. He waved a broken chair leg above his head.

'Which brings me to characterisation,' said Rabbit. 'A character should never act out of character simply to advance the plot.'"

Competing information that could distract listeners:

The information about the lessons learnt about story order, plot twists and conflict.

Listeners need to reject the lessons learnt by all the animals in favour of the only lesson that explicitly pertains to Lambkins. Lambkins' lesson about characterisation comes at the end of the text, giving it some prominence.

Test 3

Text: 'Joy Cowley', Q1:
Joy's first years at school would most likely have made her feel

Answer:
sad.

Information students need to draw on to answer this question:

"Joy found her first few years at primary school difficult."

This is the second sentence in the text and states the subject, which is explicitly expanded in the first paragraph (and is a theme of the text in general). Key information of this nature is relatively easy to remember.

Competing information that could distract listeners:

"she was good at storytelling, drawing, and making things…

'…I was a book addict…'

…By the time she reached secondary school she was writing stories for the children's section of one newspaper…

Many years later, after she had published several short stories and novels for adults…".

Listeners need to reject information about Joy Cowley having had a very successful career as a writer in favour of information about her work not being valued in her first few years at primary school. The information regarding Joy's difficult first years at school comes in the first paragraph, giving it prominence.

Test 2

Text: 'Dinosaurs in New Zealand', Q1:
Before Joan Wiffen's discovery, scientists thought

Answer:
dinosaurs had never lived in New Zealand.

Information students need to draw on to answer this question:

"For many years scientists thought that dinosaurs had never lived in New Zealand … Then, in 1979, Joan Wiffen was collecting fossils … identified it as coming from the tail of a dinosaur."

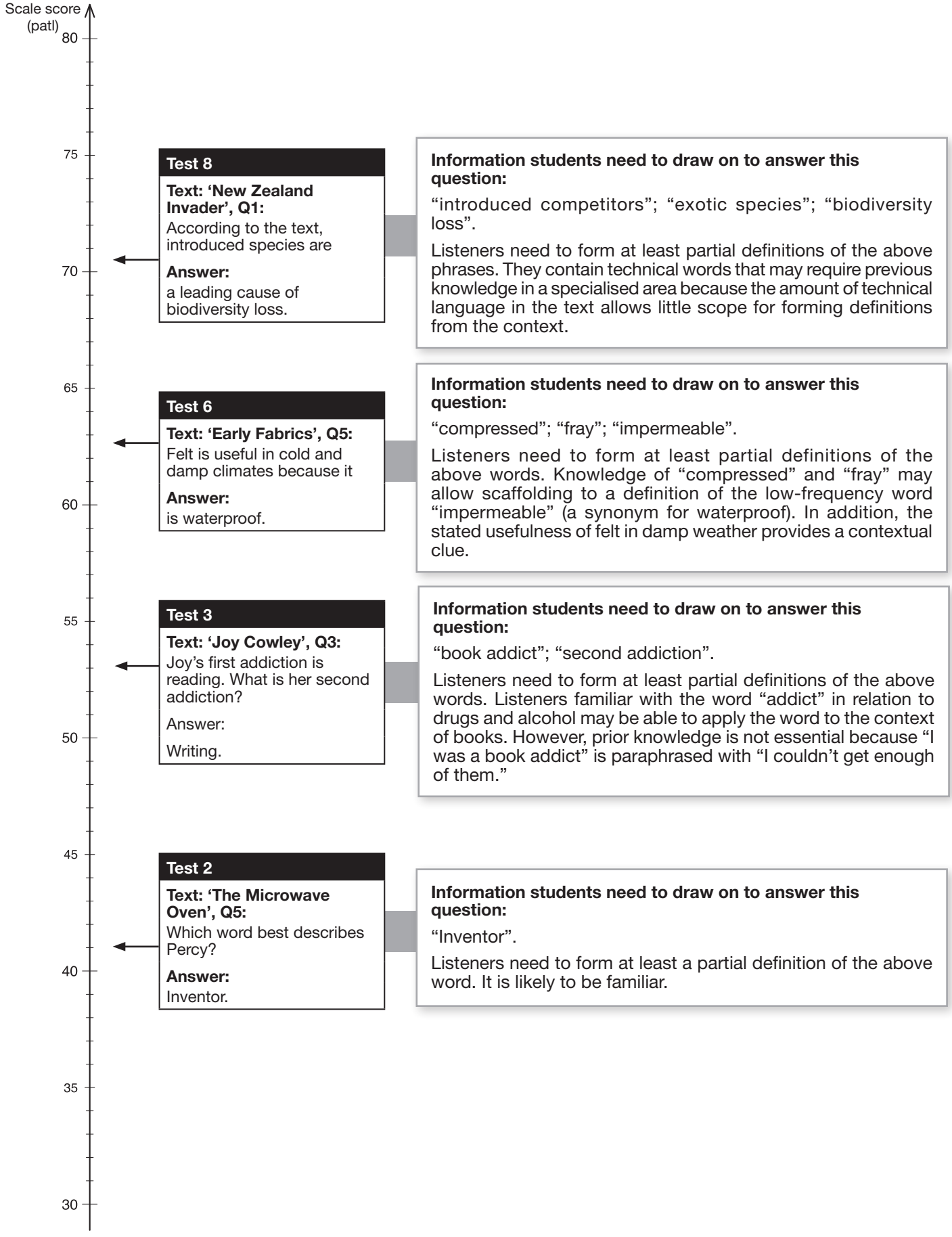
Competing information that could distract listeners:

"For a start, New Zealand's megalosaur was much smaller. Other megalosaurs were up to nine metres long and weighed over a tonne".

Listeners need to reject the specific information relating to New Zealand's megalosaur (its size, etc.) as not relevant to the mistaken judgement of scientists that dinosaurs had not lived in New Zealand. The information about the mistaken judgement comes at the beginning of the text, giving it prominence.

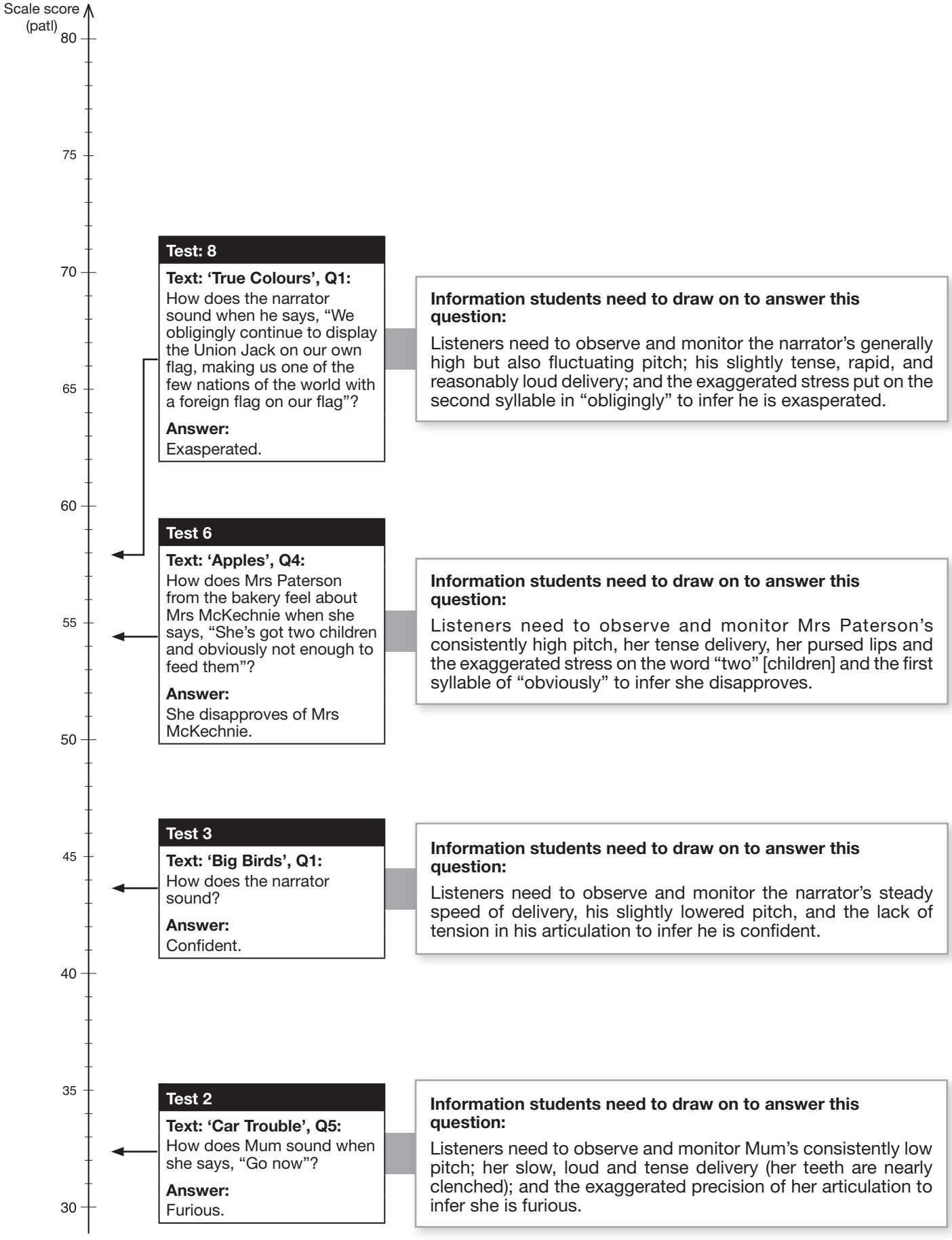
Using lexical information (vocabulary)

The examples below range from more complex vocabulary (70.3 patl) to less complex vocabulary (41.0 patl).



Using prosodic Information

The examples below range from more complex prosodic information (57.9 patl) to less complex prosodic information (32.3 patl).



References

Crooks, T., & Flockton, L. (2005). *Reading and speaking: Assessment results 2004. Report 34 prepared by the National Education Monitoring Project (NEMP) Unit*. Dunedin: National Education Monitoring Project Unit.

Crooks, T., Flockton, L., Smith, J., & Smith, L. F. (2007*). Listening and viewing: Assessment results* 2006. *Report 39 prepared by the National Education Monitoring Project (NEMP) Unit*. Dunedin: National Education Monitoring Project Unit.

Darr, C., McDowall, S., Ferral, H., Twist, J., & Watson, V. (2008). *PAT:Reading: Comprehension and Vocabulary (Revised 2008)*. Wellington: NZCER Press.

Flockton, L., & Crooks, T. (2003). *Listening and viewing: Assessment results 2002. Report 25 prepared by the National Education Monitoring Project (NEMP) Unit*. Dunedin: National Education Monitoring Project Unit.

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Kispall, A. (2008*)*. Effective teaching of inference skills for reading: Literature review. *Report prepared for the Department for Children, Schools and Families*. Slough, Berkshire, England: National Foundation for Educational Research.

Ministry of Education. (2007). *The New Zealand Curriculum*. Wellington: Author.

Ministry of Education. (2009). *The New Zealand Curriculum Reading and Writing Standards for years 1–8*. Wellington: Author.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests. Revised and expanded edition*. Chicago, IL: The University of Chicago Press. [Original work published 1960].

Reddy, M. J. (1993). The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony (Ed.). *Metaphor and thought (2nd ed.)*. Cambridge: Cambridge University Press.

Reid, N. A., Johnson, I. C., & Elley, W. B. (1994). *PAT: Listening Comprehension (Revised 1994)*. Wellington: New Zealand Council for Educational Research.

Rost, M. (2002). *Teaching and researching listening*. London, England: Pearson Education.