

# Understanding Our Test Results

A Guide for Maggidei Shiur

*Vehogisa Program — Test 4 (Mivchan Kloli) Analysis*

## What Is This Document?

This paper breaks down the results of Test 4 in the Vehogisa program. It's written for Mechanchim and Maggidei Shiur — no background in statistics or data analysis needed. Everything here is in plain language.

The test was given across three tracks (Maslulim), and we'll cover each one separately below.

**A note about test quality:** Any developing assessment program refines its questions over time — that's just how it works. In each track below, a handful of questions were flagged as needing rework. Our test-writing team is already on it. We mention them briefly so you know which scores to set aside. But the real point of this document is the teaching insights that come from the questions that ARE working well — and those insights are clear, consistent, and worth acting on.

## Background: What Is Vehogisa and How Do the Tests Work?

### The Program

Vehogisa is a Gemara learning support program running across Lubavitcher Yeshivos worldwide. The Year 1 goal is simple: by the end of Beis Mesivta, every bochur should be able to independently learn an Amud Gemara with Rashi.

To get there, the program provides Dappei Ezer (guided worksheets) that walk bochurim through the Gemara, building key skills along the way. Regular tests track progress — but these tests aren't just about checking how much a bochur remembers.

### The Purpose of Testing

The test is *not* designed to show how much a bochur knows and remembers. Instead, it's designed so that the **Maggid Shiur and the bochur together** can see which skills the student

excels in and which need more work. That way, we can tailor the learning and zero in on exactly where the talmid needs support.

After each test, our team of Mechanchim and curriculum experts sit together and go through the results. They figure out what's working and what needs more focus. Then the Maggidei Shiur adjust their approach, and the test writers start building the next assessment. Teach, test, learn from the data, improve. That's the cycle.

## The Three Core Skills We Test

Every question on the test measures one of three core skills a bochur needs for independent Gemara learning:

**Skill 1 – Recall (Vocabulary and Definitions):** Can the bochur understand the basic language and terminology of the Sugya? Does he know what key terms mean and what principles they refer to? This is the most fundamental skill — without it, a bochur simply can't read the Gemara.

**Skill 2 – Reasoning (Understanding and Logic):** Can the bochur go beyond definitions and actually grasp the reasoning behind what he's learning? Can he follow the Gemara's logic — why Chazal said what they said, how one point leads to the next? This takes more than memory. It takes real comprehension.

**Skill 3 – Inference (Application to New Situations):** This is the hardest one. Can the bochur take what he learned in the Sugya and apply it to a situation he hasn't seen before? Can he read between the lines and figure out what's implied? This is the skill that ultimately lets a bochur learn on his own.

Every test question falls into one of these three categories. So when we look at results, we don't just see an overall score — we can see which specific skills a bochur is strong in and which need work.

## The Three Tracks (Maslulim)

Bochurim are placed into one of three tracks, based on how much material they're covering:

**Maslul Aleph** – Covers the fewest Dafim, allowing more time per Daf, often learned entirely in Seder.

**Maslul Beis** – Middle coverage.

**Maslul Gimel** – The most Dafim. Often learned in part out of Seder.

Each track gets its own version of the test, covering whatever material that track has learned. The skill areas (Recall, Reasoning, Inference) and question difficulty are the same across all tracks — the only difference is which Dafim the questions draw from.

## How to Read This Analysis

When we look at test results, we're really looking through two lenses. Both matter, but one is far more relevant to your day-to-day work in the classroom.

### **Your Primary Lens: What Are the Bochorim Struggling With?**

When a solid question is being missed by lots of bochorim, that's a direct signal about a real skill gap that needs attention in the classroom. This is the information that makes your teaching more effective — and it's the heart of this document. The analysis below focuses mainly on these teaching insights.

### **The Secondary Lens: Are Any Questions Not Working Properly?**

Before we draw conclusions about bochorim from any specific question, we need to make sure the question itself is doing its job. A small number of questions in each track had issues — like strong students getting a question wrong more often than weak students (which tells you the question is flawed, not the students). We note these briefly in each section so you know which scores to ignore. The test-writing team is handling the fixes.

**The key principle:** Once we set aside the handful of questions that need rework, the rest of the results paint a clear and consistent picture of where bochorim are strong and where they need help. That picture is what this document is really about.

## Maslul Aleph (Fewest Dafim)

### What the Results Tell Us About the Bochurim

|           | Average Score | Middle Score |
|-----------|---------------|--------------|
| Recall    | 82%           | 80%          |
| Reasoning | 68%           | 60%          |

#### The Core Finding: Recall Is Solid — Reasoning Is the Growth Area

Recall averages 82%, Reasoning averages 68%. That 14-point gap? It's the single most important teaching insight from this test.

Most bochurim in Maslul Aleph can handle the basic vocabulary and definitions — they can read the Gemara and understand the terms. But a lot of them are struggling to explain WHY the Gemara says what it says, or how one point connects to the next.

The wider spread in Reasoning scores tells us something else: some bochurim are very strong in reasoning while many others are still developing it. This isn't a "everyone is weak" situation — reasoning ability varies a lot, which means targeted, individualized attention is going to make the biggest difference.

#### Are the Two Skills Connected?

There's a moderate connection. Bochurim who are strong in Recall tend to be somewhat stronger in Reasoning too — which makes sense. You need to understand the words before you can reason about them.

**The key classroom takeaway:** Just because a bochur can define terms and recall facts does NOT mean he can follow the logic. If a bochur seems to "know the material" but can't explain the reasoning, that's a specific, identifiable skill gap. It's not that he "isn't trying hard enough" — he needs focused practice on a different type of thinking.

#### Questions We Can Trust (Reliable Anchors)

**Strong Recall questions:** Q3, Q20, Q23 — these consistently separate stronger from weaker students.

**Strong Reasoning questions:** Q5, Q7 — especially effective at showing who truly understands the logic.

These are the questions our skill insights above are built on. They're also good models for writing new questions.

#### A Note on Test Quality

*The following is included for transparency. The test-writing team is already addressing these issues.*

A few questions in Maslul Aleph aren't working as intended yet. This is normal in a developing program and doesn't change the core teaching insights above, which are based on the reliable questions.

| Question | Issue                                       | What's Being Done                  |
|----------|---|------------------------------------|
| Q22, Q24 | Not aligning with their skill category      | Answer key / wording being audited |
| Q4, Q25  | May be testing Recall rather than Reasoning | Under review for reclassification  |
| Q6       | 100% correct — too easy to measure anything | Will be kept as warmup only        |

## Action Items for Maslul Aleph

### For the Maggidei Shiur

1. **The data shows that Recall is landing well** — so keep doing what you're doing for vocabulary and definitions.
2. **Invest significantly more time on reasoning:** not just “what does it say” but “why does it say it” and “how does this connect to what came before?”
3. For bochurim who score well in Recall but poorly in Reasoning — that's a specific skill gap. It won't fix itself with more of the same learning. They need targeted practice.

### For the Test Writers

1. Audit and fix Q22 and Q24 first — biggest impact on test quality.
2. Review Q4 and Q25 — clarify wording or reclassify.
3. Decide on Q6 — keep as warmup but remove from scored skill total.
4. Target 8–10 questions per skill area in future tests for stronger measurement.

## Maslul Beis (Middle Coverage)

### What the Results Tell Us About the Bochorim

#### The Core Finding: Recall Is the Foundation of Everything

Recall has a very strong connection to the General Score — bochorim who do well on Recall tend to do well overall. It's the strongest relationship of any skill section on this test.

Here's what that means for teaching: if a bochor's Recall is weak, everything else is going to suffer. Vocabulary and definitions are the bedrock. Make sure every bochor has these down solid before pushing into higher-level skills.

Reasoning has a moderate-to-strong connection to overall success. Inference currently shows almost no connection — but that's because of a few broken questions (especially Q10), not because Inference doesn't matter. Once those questions are fixed, we expect this picture to change.

### How Well Is Each Skill Section Working?

| Skill Section | Reliability | What This Means   |
|---------------|-------------|---|
| Recall        | Strong      | These questions work well together — you can trust this score.                    |
| Reasoning     | Weak        | Some questions might be testing a different skill — take it with a grain of salt. |
| Inference     | Weak        | At least one question (Q10) is badly broken and dragging the whole section down.  |

**Bottom line:** Trust the total score and the Recall subscore. But take the Reasoning and Inference subscores with a grain of salt until the problem questions get fixed.

#### Questions We Can Trust (Reliable Anchors)

**Reasoning Q22** (40% correct, excellent at separating students), **Recall Q6 & Q17** (solid discriminators), **Inference Q13 & Q16** (effective application questions).

#### A Note on Test Quality

*The following is included for transparency. The test-writing team is already addressing these issues.*

**Highest priority — Q10 (Inference):** The strongest students got Q10 right only 20% of the time, while the weakest students got it right 60% of the time. That's the opposite of what should happen — a clear sign the problem is with the question, not the students. Usually this means an incorrect answer key or confusing wording. Just removing this one question alone would significantly improve the quality of the entire test.

**Q2 (Reasoning):** 80% got it right, but it's not separating strong from weak students — it's too straightforward to really work as a Reasoning question.

**Q4 (Inference):** 50% correct, but also not separating strong from weak. Probably a wording or answer key issue.

**Questions being reclassified:** Q8 (Reasoning → Inference), Q4 (Inference → Reasoning), Q21 (Reasoning → Recall). These aren't bad questions — they just need to land in the right category so the skill scores are accurate.

## Action Items for Maslul Beis

### For the Maggidei Shiur

1. **Recall is the foundation.** If bochurim are weak here, prioritize vocabulary and definitions before anything else.
2. **Once Recall is solid, begin building Reasoning skills** — the “why” behind the “what.”
3. Don't draw conclusions from individual bochur Inference scores just yet — the test questions in that area need fixing first.

### For the Test Writers

1. Fix or replace Q10 first — this single change will have the biggest impact.
2. Audit Q2 and Q4 — check answer keys and wording.
3. Reclassify Q8, Q4, and Q21 into their correct skill categories.
4. Add stronger Reasoning and Inference questions — with only 5 per section, even one weak question drags everything down.

## Maslul Gimel (Most Dafim)

### What the Results Tell Us About the Bochurim

Good news here: the test as a whole is quite coherent. All three skill sections connect strongly with the General Score, which means the test IS measuring real ability.

| Skill Section | Connection to General Score | Strength           |
|---------------|-----------------------------|--------------------|
| Recall        | 91%                         | Very Strong        |
| Reasoning     | 94%                         | Very Strong (Best) |
| Inference     | 91%                         | Very Strong        |

### The Core Finding: Reasoning Is the Key to Overall Success at This Level

In Maslul Gimel, Reasoning is the single strongest predictor of overall performance. If you could only focus on one area to move the needle on a bochur's overall growth, this would be it.

This is a big deal: for bochurim at this level, the ability to understand WHY the Gemara says what it says — not just WHAT it says — is what separates the ones who are progressing from the ones who are stuck.

All three skill areas are close behind each other, so it's not about neglecting Recall or Inference. But when a bochur at this level starts to grasp reasoning, everything else tends to follow.

### One Important Skill Insight

**Q19:** Almost everyone missed this one, but the stronger bochurim were still more likely to get it right. So the question is working fine — it's just really hard. **This is a teaching signal, not a test problem.** Whatever skill Q19 is testing, most bochurim at this level haven't developed it yet. Worth looking at what concept this question covers and whether it needs more classroom time.

### A Note on Test Quality

*The following is included for transparency. The test-writing team is already addressing these issues.*

**Red flags — Q5, Q18, Q21:** These three questions show “negative discrimination” — bochurim with higher overall performance were NOT more likely to get them right. That's a clear sign the problem is with the question. Do NOT use these three results to draw any conclusions about individual bochurim's skills.

**Too easy — Q1, Q12:** Every single bochur got these right. Fine as confidence-builders, but they give us zero measurement information.

**Being reclassified:** Q15 (Reasoning → Recall), Q3 & Q6 (Reasoning → Inference), Q16 (Inference → Reasoning), Q21 (Inference → Recall), Q18 (Recall → Reasoning). Note: Q21 and Q18 are also red-flag questions — fix them first, then reclassify.

## Action Items for Maslul Gimel

### For the Maggidei Shiur

1. **Reasoning is the key to unlocking overall success.** Help bochurim understand the WHY, not just the WHAT.
2. Look at Q19 results for individual bochurim — this is a real skill gap that most haven't bridged yet.
3. Don't draw conclusions about individual bochurim from Q5, Q18, or Q21 scores — these questions aren't reliable and need fixing first.

### For the Test Writers

1. Audit Q5, Q18, and Q21 right away — check answer keys and wording.
2. Replace Q1 and Q12 with slightly harder versions so we actually get useful data.
3. Reclassify questions: start with Q15, Q3, Q6, Q16 (strongest evidence).

## Summary: The Big Picture Across All Three Tracks

### Common Themes

- 1. Recall is the foundation.** Across all tracks, bochurim who are strong in basic vocabulary and definitions tend to do better overall. The Dappei Ezer's focus on language skills is the right starting point.
- 2. Reasoning is where the real growth happens.** In Maslul Aleph, the 14-point gap between Recall (82%) and Reasoning (68%) shows this is the next frontier. In Maslul Gimel, Reasoning is actually the best predictor of overall success. Getting bochurim from "I know what it says" to "I understand why it says it" — that's the central challenge.
- 3. A handful of test questions need fixing — and that's normal.** Every track has a few questions that aren't working properly. Totally expected in a developing program. The key thing: don't draw conclusions about bochurim's abilities from those questions. The test team is on it.
- 4. The working questions paint a clear picture.** Even after setting aside the questions that need rework, we've got a strong body of reliable data. The patterns are consistent across tracks and directly useful in your classroom.

### An Interesting Observation Across Tracks

The questions across all three tracks test the same skill areas at the same difficulty level — the only difference is which Dafim they cover. But when we compare the results, something interesting jumps out.

Maslul Gimel (the boys covering the most Dafim) actually shows the strongest, most consistent performance across the board. Meanwhile, Maslul Aleph (the boys covering the fewest Dafim) shows a noticeable gap between Recall and Reasoning.

This suggests that the bochurim in Maslul Gimel tend to be stronger, more independent learners to begin with — they've already got solid foundational skills and can handle a faster pace. The bochurim in Maslul Aleph likely need the extra time precisely because they're still building core skills, especially Reasoning.

That's valuable context for Mechanchim: the track a bochur is in actually tells you something about where he probably needs the most support.

### Quick Reference: How to Use Test Results in Your Classroom

| What You See in a Bochur's Results | What to Do   |
|------------------------------------|--|
| Weak in Recall                     | Focus on the basics: vocabulary, definitions, understanding what the words mean. The Dappei Ezer worksheets are built for exactly this. Make sure he's working through them carefully before moving on.                              |
| Strong Recall, Weak Reasoning      | He knows the words but isn't seeing the logic. Ask more "WHY" questions: Why does the Gemara ask this? Why does Rashi explain it this way? How does this connect to what came before? Help him see the thread tying it all together. |

|  |  |
|--|--|
| <p>Strong Recall &amp; Reasoning, Weak Inference</p> | <p>He understands what he's learned but can't apply it to new situations. Give him scenarios: "If the Gemara says X about this case, what would it say about a different case?" This is the skill that eventually makes independent learning possible.</p> |
| <p>Low score on a flagged question</p>               | <p>Just ignore it. If a question's been flagged as a test-side problem, that score doesn't tell you anything about the bochur's skill. Wait for the fix before drawing conclusions.</p>  |

Remember: the point of these grades isn't to rank bochurim. The point is to give YOU the information you need to help each bochur grow in the specific areas where he needs it most. Every bochur has strengths. Every bochur has areas that need work. This analysis is here to help you see both clearly — and teach accordingly.

Vehogisa — Igud Yeshivos Lubavitch  
[igudyeshivos.com/vehogisa](http://igudyeshivos.com/vehogisa)