The 2014 GED® Test - Science

Short Answer Resource Guide for Adult Educators

March 2014

The 2014 GED® Test - Science
Short Answer Resource Guide for Adult Educators

Table of Contents

Overview and Introduction to GED® Test - Science Short Answer Resource Materials

This guide has been assembled by GED Testing Service in order to help adult educators increase their understanding of and skill in scoring the Short Answer (SA) questions that appear on the 2014 GED® test - Science. Using these resources will help you identify the various qualities and attributes of SA responses at the full range of score points on scoring guides relating to four example Science SA items. Each SA item that appears on the operational GED® test or on GED Ready™: The Official Practice Test has a unique scoring guide applicable to that particular item only. However, studying and working with the materials in this resource guide will enable you to develop an understanding of how these scoring guides are applied to a variety of responses. Further, this guide may help you focus instruction in scientific thinking skills and response techniques to SA items in general.

Using these materials will also help you in scoring SA responses that adult learners provide you as part of their preparation for the test in taking GED Ready™: The Official Practice Test. GED Ready™ is accompanied by the Educator Scoring Tool for GED Ready™, available on the 2014 test resources webpage, which can help you score test-taker responses. There is one tool per SA item that appears in GED Ready™. This guide, as a supplement to these tools, is intended to increase your facility with and accuracy in scoring SA items for the 2014 GED® test - Science.*

The materials in this guide are based on two items that appear on GED Ready™ and two items that appear on the free practice test. The first question and its accompanying exemplar responses are based on a textual stimulus material that asks test-takers to examine the relationship between wind energy use and the energy supply of coal. The second question asks test-taker to design an investigation exploring solubility. The third question examines the relationship between a particular species of ants (*Ophiocordyceps unilateralis*) and a species of fungus. The fourth question asks test-takers to design an experiment to test a hypothesis about soil erosion. The stimulus materials and associated prompts were part of an extensive field-testing process that each of the questions on the 2014 GED® test underwent in 2012. Stimulus materials and prompts for questions from GED Ready™ are available online; stimulus materials and prompts for questions from the free practice test are incorporated into this guide. The answers to these questions that you will see in this guide are actual responses written by adult test-takers during that field test. These exemplar responses were written under standardized computer-based testing administration conditions that replicate the conditions of actual operational GED® testing on computer in all respects (e.g., instructions provided to test-takers, tools available to test-takers, time allotment, etc., were identical to authentic testing conditions). All of the characteristics of the responses, including spelling, paragraphing and spacing, have been left exactly as originally written and submitted by the test-takers. They also appear here exactly as they appeared to the educator Subject Matter Experts (SMEs) who determined the range of responses for each score point and to the expert human scorers who provided the final certified scores for the responses. The annotations that are presented to enhance your understanding of the score each response received were also written by SMEs.

* Note: Each SA scoring tool is meant to be used as a guide to scoring, but once you become more familiar with the manner in which the SA rubrics are constructed, you will likely be able to score SA response samples without fully following the tool. There is no expectation that you will use the tool for EVERY response that you score, and the materials in this guide should help you begin to gain the skills at evaluation of writing that you will need to effectively score SA responses first with the tool and, later, without relying on it.

GED Ready™: The Official Practice Test – Science

The source materials and the prompt for the *Wind Energy* SA item that appears on GED Ready™ – Science can be downloaded at http://www.gedtestingservice.com/stimulus-opt-in.

Wind Energy Science Short Answer Scoring Guide

Question Overview: In this SA question, test-takers are required to give an explanation about how a significant increase in the used of wind energy would affect the energy supply of coal. Then, they must also cite specific evidence from the data table to support their reasoning as to why wind energy would be a preferred energy source over coal. This question tests learners' skill at the complex task of using, producing, and justifying a text-based line of reasoning by incorporating elements from the data table into the presentation of their own ideas.

Scoring Guide: Each response is scored on the basis of two key elements. Each bullet below describes the *quality* of these elements typical of each score point.

3-Point Response

Response contains

- a clear and well-developed explanation of how a significant increase in the use of wind energy would affect the energy supply of coal
- complete support from the data table

2-Point Response

Response contains

- an adequate or partially articulated explanation of how a significant increase in the use of wind energy would affect the energy supply of coal
- partial support from the data table

1-Point Response

Response contains

- a minimal or implied explanation of how a significant increase in the use of wind energy would affect the energy supply of coal
- minimal or no support from the data table

0-Point Response

Response Includes

- no explanation of how a significant increase in the use of wind energy would affect the energy supply of coal
- no support from the data table

Non-scorable Responses (Score of 0/Condition Codes)

Response exclusively contains text copied from source text(s) or prompt
Response demonstrates that the that test-taker has read neither the prompt nor the source text(s)
Response is incomprehensible
Response is not in English
Response has not been attempted (blank)

Science Short Answer Responses and Annotations for *Wind Energy*

Text from the responses quoted within the annotations is highlighted in yellow in both the annotations and the test-taker response to help you quickly identify specific elements of each response that helped SMEs score them appropriately. However, keep in mind that each response must be considered as a whole, and these highlighted excerpts are notable mostly because they show specific examples of qualities common to responses.

Test-taker Anchor Response 1 – Score: 3	**Annotation**
1) The most important piece of data that wind energy versus coal energy is better for all would be the difference in CO2 emissions. Coal releases about 200 lbs per kWh, where as wind produces none at all. 2) The fact that coal has an ongoing fuel cost and wind does not, would make wind energy a better choice of the two. 3) Though wind has some negitive points, ex: visual and noise, these are very minimal compared to coal. The long term effect of the contamination of burning coal are non-reversable. The current supply of coal only has a 100 year life span. This is a very Short time in the over all realm of things. If would could increase the use of wind energy to absorb at leat 10% of the 40% of coal that is used for electricity we would be able to reserve our supply of coal to last us longer. Not only that, but the cost to produce that extra 10% would electricity though wind would be a significate savings of funds. We would also be saving on the fact that the less coal we are converting the less contamination we are producing and having to clean up. The more wind energy we produce the less we have to draw on our coal supply. This will result in our coal supply to exstead to a longer life span.	Score Explanation: The test-taker response includes complete support from the table by explaining the preference for wind energy, with a direct comparison (Coal release about 200 lbs per kWh, where as wind produces none at all) and (coal has ongoing fuel cost and wind does not). The third paragraph also provides a qualitative comparison based on table data. The test-taker's well-developed explanation for the impact on the coal energy supply with increased wind power use is detailed and clear, (more wind energy we produce the less we have to draw on our coal supply. This will result in our coal supply to exstead to a longer life span). Holistically, this response contains well-developed ideas throughout, which are typical of responses that receive scores of 3.

Test-taker Anchor Response 2 – Score: 3	Annotation
Wind energy would be a preferred energy source over coal because: Wind releases no co2 emissions. Wind has no ongoing fuel cost. Wind is renewable. Wind releases no airborn contaminations. Wind does not require strip mining. Wind does not contaminate groundwater. The same cannot be said for coal for any of the above listed. A significant increase in the use of wind energy would keep the coal supply from being used up so quickly, thereby insuring the coal supply will last longer.	Score Explanation: This response contains a concise comparison with strong support in the form of multiple citations from the table for why wind energy would be preferable to coal. For example, solid table interpretations are used to explain a preference for wind in view of the negative aspects of coal, *(*Wind has no ongoing fuel cost…Wind releases no airborn contaminations…Wind does not require strip mining.). The test-taker also provides a well-developed explanation for the impact of increased use of wind energy on the coal energy supply, (keep the coal supply from being used up so quickly, thereby insuring the coal supply will last longer*)*. Although the explanation lacks comparisons from the data table, the response contains a well developed explanation for the impact of increased use of wind energy, which helps compensate for the lack of detail in support from the data table. Therefore, this response receives a score of 3.

Test-taker Anchor Response 3 – Score: 3	Annotation
Wind energy is far more resourceful compared to coal. Coal has ongoing fuel costs, and has many more impacts to the environment as wind energy does not. Wind does not produce any Co2 emissions, as where coal produces 200 pounds of CO2 emissions per kWh. Coal has many impacts that will affect the environment in the long run, as to wind energy. simple inconveniences are the main issue. The effects of wind energy would help with the coal consumption. Projections show that coal resources will only last 100 years if we don't find alternate energy sources.	Score Explanation: This test-takers response includes complete support from the table. (Coal has ongoing fuel costs , and has many more impacts to the environment as wind energy does not) and (Wind does not produce any Co2 emissions, as where coal produces 200 pounds…). Additionally, the test-taker provides a holistically complete explanation for how the energy supply of coal will be affected, (wind energy would help coal consumption. Projections show that coal resources will only last 100 years if we don't find alternate energy sources.). Although the explanation lacks some of the development that is typical for responses that receive scores of 3, this particular response uses high quality support from the data table, which helps compensate for the lack of detail in the explanation.

Test-taker Anchor Response 4 – Score: 2	Annotation
Wind energy does not produce any carbon dioxide. Although the amount of energy produced by wind depends on wind speed, that energy can be banked for future use. There are no ongoing fuel costs with wind. Increasing the use of wind energy would reduce the consumption of the limited supply of coal.	Score Explanation: The test-taker provides implied comparisons for why wind is preferable. These implied comparisons translate into partial support from the table, (Wind energy does not produce any carbon dioxide.) (accepting this as emissions) and (no ongoing fuel costs with wind). The test-taker states that the coal supply is *(*limited*)*, so the inference is that coal is non-renewable. The test-taker also includes a clear explanation for how a significant increase in the use of wind energy would affect the coal supply, *(*Increasing the use of wind energy would reduce the consumption of the limited supply of coal*)*. Although the response shows some weaknesses in the comparisons from the data table, the clarity of the explanation for how the increase in wind energy would affect the coal supply is so strong that this response receives a holistic score of 2.

Test-taker Anchor Response 5 – Score: 2	Annotation
The energy from wind seems to be a preferred energy source from looking at the data presented in the table. The carbon dioxide emissions are listed as none and the ongoing fuel costs seems to none as well. So it looks like the wind as a energy source would be preferable. Looking at the data for the energy from coal the carbon dioxide emissions seems to be very high about 200 pounds. The data also shows other impacts from coal energy as water contamination, mercury contamination and as a non-renewable fuel source. If wind energy is used primarily as a energy source the coal supply will be affected.	Score Explanation: The test-taker completely supports the explanation for why wind energy would be preferred. The response provides clear comparisons between coal (…carbon dioxide emissions seems to be very high about 200 pounds. The data also shows other impacts from coal energy as water contamination, mercury contamination and as a non-renewable fuel source.) and wind (The carbon dioxide emissions are listed as none and the ongoing fuel costs seems to none as well.). The test-takers use of the word "*none*" imparts a qualitative interpretation of the table data that is acceptable. However, the test-taker's explanation for the impact on the energy supply of coal is insufficient, (…will be affected). While the explanation for the impact on the energy supply is insufficient, the clear and well-developed evidence from the data table strengthens the response, completely supporting the explanation for why wind energy is a preferred source. Therefore, this response receives a score of 2.

Test-taker Anchor Response 6 – Score: 2	Annotation
Wind energy is better because it has zero emissions and has no ongoing fuels cost.coal is nonrenewable so one can be able to save coal only when needed plus coal mining is toxic to the environment.	Score Explanation: The test-taker provides a partially supported explanation for the preference of wind energy by using support from the table, (zero emissions) (CO_2 emissions) and *(*no ongoing fuel cost*)*. The statement that (coal is non-renewable) provides an acceptable inference that wind is renewable (not stated in the table). However, the attempt to explain the effect on the energy supply of coal is very weak, *(*save coal only when needed*)*. This explanation lacks the development typical for responses that receive a score point of 2. However, the response includes adequate support from the data table, which helps compensate for the lack of clarity in the explanation.

Test-taker Anchor Response 7 – Score: 1	Annotation
The steam drives a turbine, which turns an electric generator. the increase of wind use would affect the energy supply of coal cause it would lower ongoin fuel cost also it would cut down on airbourne mercury contamination in the air.	Score Explanation: This test-taker's response provides a partially-supported comparison by using information from the table, (…lower the ongoin fuel cost also it would cut down on airbourne mercury…). The explanation of how the use of wind energy would affect the coal supply is insufficient, (…would affect the energy supply of coal…). Holistically, the response is very weak in that all of the test-taker's ideas are underdeveloped. Therefore, the response receives a score of 1.

Test-taker Anchor Response 8 – Score: 1	Annotation
With wind Co2 Emissions per kWh at zero and coal at around 200 pounds. it evidently shows that this is much more effective in that particular area. The power availability for coal 24x7, 365 days per year and wind, which varies directly with wind speed doesn't sound to promising even though the ongoing fuel costs for the coal is apparent and not for wind because of it's natural recourse, I wouldn't invest my money into a natural source that isn't consistent. Other significant negative impacts that would take place with coal in regards to stip mining & groundwater contamination, airborne mercury contamination can easily be controlled with chemicals to treat both to purify the water and the air. The wind, on the other hand, could impose convictions with the animal rights activists, simply because it poses a danger to birds. The large obstruction of scenery can be a burden to nearby residents in the general area. The noise issues as well might burden the surrounding residence and or business (if they reside in the area) In conclusions, if conversions from coal to wind were to come about in the future, it would rule out coal in the long run. Though, I believe coal is a lot cheaper than purchasing wind turbines and the variations that go along with it from China when we have available coal right here in the northern parts of the US, it would be a waste to spend a significant amount of money importing from China vs. using what we already have here. This would also create jobs in America thus, lowering our unemployment rate. If we don't utilize the coal, it will just sit there as we focus on a insignificant source if the wind weren't to blow.	Score Explanation: This response gives support from the table to compare the differences between coal and wind energy (With wind CO_2 Emissions per kWh at zero and coal at around 200 pounds…power availability for coal 24x7, 365 days per year…), but the support is partial. The explanation takes the form of an evaluation as one source is not preferentially selected. Additionally, this response does not sufficiently explain the effect of using wind power on the energy supply of coal, (…it will just sit there…). Overall, this response receives a score of 1. The test-taker's ideas are vague and underdeveloped and they fail to provide a clear preference for why one energy source would be preferred over another.

Test-taker Anchor Response 9 – Score: 1	Annotation
Point #1 There is only enough coal for about another 100years. Point#2 As renewable energy increases over time the need for coal will decrease. Point#3 By the use of renewable energy over time it will reduce the carbons in the air. Then hopefully with less toxins in the air we should have better oxygen rates.	Score Explanation: This response does not provide support from the table to compare coal and wind. The test-taker gives an adequate explanation for the effect on the energy supply of coal, (…over time the need for coal will decrease). Although the response completely lacks the support from the data table that is typical of responses that receive scores of 1, this is compensated for by an adequate explanation for the effect on the energy supply of coal. Therefore, it earns a score of 1.

Test-taker Anchor Response 10 – Score: 0	Annotation
When we burn up coal it is toxic to us. It is costing a lot, weighs a lot, needs to be worked on more. If we used wind energy, it would be cheaper, safer, lighter.	Score Explanation: The test-taker compares the two sources and makes a case for for why wind energy (…cheaper, safer, lighter) is better than coal (…toxic…costing a lot, weighs a lot…), but it is unclear whether the information was interpreted from the table. The test-taker does not attempt to explain the effect on the energy supply of coal. Although this response does include some reasons for why wind energy is better than coal that are typical of responses that receive scores of 1, the comparisons do not appear to be taken from the data table. Additionally, there is no attempt to provide any explanation as to how the use of wind energy would affect the coal supply, therefore this response is too weak to earn a score.

Test-taker Anchor Response 11 – Score: 0	Annotation
Wind energy does not have major negative impacts. and any cons it does have. Are with in a reasonable solutions. Wind energy or renewable energy is a supply in demand, which in a couple of years it will play a big roll in the way we generate electricity, and mayby affect job placement with coal companys and power plants.	Score Explanation: This response identifies wind energy as (…renewable…), which could be inferred from either the table or the text. However, the test-taker fails to provide a comparison for why wind energy would be preferred to coal or another source. This response does not address the energy supply; therefore, the overall response is too weak to earn a score.

Test-taker Anchor Response 12 – Score: 0	Annotation
CO2 Emissions – Zero Power availability -24/7.365 days per year NO fuel costs	Score Explanation: This response provides information from the data table from different categories with no identification or explanation. This test-taker does not attempt to explain any change in the energy supply of coal. Therefore, the response is too weak to earn a score higher than zero.

GED Ready™: The Official Practice Test – Science

The source text and the prompt for the *Solubility Experimental Design* SA item that appears on GED Ready™ – Science can be downloaded at http://www.gedtestingservice.com/stimulus-opt-in.

Solubility Experiment Design Science Short Answer Scoring Guide

Question Overview: The purpose of this SA question is to provide the test-taker with an opportunity to design a scientific investigation. The test-taker must understand and apply the fundamentals of scientific investigation design in his or her response. In addition, the test-taker must describe an experimental set-up to test a specific hypothesis, which includes a description of the methods used to collect data. The test-taker must also justify his or her line of reasoning used to determine whether the hypothesis is valid. Correctly completing the multi-step process required by this question involves several critical thinking skills, including identifying a research question, designing a scientific investigation, and justifying a line of reasoning.

Scoring Guide: Each response is scored on the basis of three key elements. Each bullet below describes the *quality* of these elements typical of each score point.

3- Point Response

Response contains

- a well-formulated, complete controlled experimental design
- a well-formulated data collection method
- a well-formulated, complete explanation of the criteria for evaluating the hypothesis

2-Point Response

Response contains

- a logical controlled experimental design
- a logical data collection method
- a logical explanation of the criteria for evaluating the hypothesis

1-Point Response

Response contains

- a minimal experimental design
- a minimal or poorly-formulated data collection method
- a minimal or poorly-formulated explanation of the criteria for evaluating the hypothesis

0-Point Response

Response contains

- an illogical or no experimental design
- an illogical or no data collection method
- an illogical or no explanation of the criteria for evaluating the hypothesis

Non-scorable Responses (Score of 0/Condition Codes)

Response exclusively contains text copied from source text(s) or prompt
Response demonstrates that the that test-taker has read neither the prompt nor the source text(s)
Response is incomprehensible
Response is not in English
Response has not been attempted (blank)

Science Short Answer Responses and Annotations for *Solubility Experimental Design*

Text from the responses quoted within the annotations is highlighted in yellow in both the annotations and the test-taker response to help you quickly identify specific elements of each response that helped SMEs score them appropriately. However, keep in mind that each response must be considered as a whole, and these highlighted excerpts are notable mostly because they show specific examples of qualities common to responses.

Test-taker Anchor Response 1 – Score: 3	**Annotation**
Experimental setup: Prepare fifteen beakers with identical masses of distilled water. Ensure that these beakers are of identical dimensions. Three trials will be completed, with five different beakers used for each trial. Label the beakers with letters to correspond with the trial (a, b, or c) and temperature (1, 2, 3, 4, or 5). Using the balance, prepare a number of samples of salt of identical, small mass. Procedure for data collection: Perform the three trials concurrently or in immediate succession. For each trial, record the amount of salt, by mass, that dissolves completely (before becomming supersaturated) in beaker 1 (at room temperature). Then heat beaker 2 to 150% of the room temperature. Record the amount of salt that it dissolves. Heat beaker 3 to twice room temperature and record the amount of salt that it dissolves. Heat beakers 4 and 5 to three and five times room temperature, respectively, and record the amount of salt that each dissolves. Evaluation: Tabulate the data, with the temperature arranged horizontally and the amount of dissolved salt vertically. A scatterplot, with temperature on the horizontal axis and mass of dissolved salt on the vertical, will be useful for interpretation of the data. If each of the trials shows an increase in dissolved salt as temperature increases, then the experiment has confirmed the hypothesis. The degree of consistency between the results of the different trials will speak to their reliability.	Score Explanation: The test-taker's response includes a logical experimental design demonstrating use of a manipulated variable, temperature, (…fifteen beakers with identical masses of distilled water…identical dimensions…Three trials…beaker 1 (at room temperature)…heat beaker 2 to 150% of the room temperature…Heat beaker 3 to twice room temperature…Heat beakers 4 and 5 to three and five times room temperature…record the amount of salt that each dissolves.) The response has a well-formulated data collection method (. . . record the amount of salt, by mass, that dissolves completely [before becoming supersaturated*).* The third paragraph also includes a complete explanation of the criteria for evaluating the hypothesis that connects the collected data back to the hypothesis (Tabulate the data, with the temperature arranged horizontally and the amount of dissolved salt vertically. ..If each of the trials shows an increase in dissolved salt as the temperature increases, then the experiment has confirmed the hypothesis). Overall, this response contains well-developed ideas throughout, typical of responses that receive scores of 3.

Test-taker Anchor Response 2 – Score: 3	Annotation
Allow for three beakers to be filled with equal amounts of water. The first beaker, acting as the control group, will be left at room temperature and have salt added in increments until no more salt can be dissolved, measuring each amount with the balance. Record the amount of salt dissolved into the first beaker, in grams. The second beaker will be heater to 100F, and similarly, salt will be added in increments until no more can be dissolved. Record the amount of salt added in grams. The third beaker will be heated to boiling, and salt will be added in increments until no more can be dissolved. Record the amount of salt added in grams. Should more salt be dissolved in each beaker as their temperature increases, it would validate the hypothesis.	Score Explanation: This response contains a description of a logical experimental design demonstrating use of a manipulated variable, temperature, (…equal amounts of water…first beaker, acting as the control group, will be left at room temperature and have salt added in increments until no more salt can be dissolved…second beaker will be heated to 100F, and similarly, salt will be added in increments…The third beaker will be heated to boiling.) There is a well-formulated data collection method that provides a clear description of how the amount of remaining salt will be precisely measured (. . . salt added in increments until no more salt can be dissolved, measuring each amount with a balance). Finally, the test-taker includes a complete explanation of the criteria for evaluating the hypothesis, which connects the collected data back to the hypothesis (Should more salt be dissolved in each beaker as their temperature increases, it would validate the hypothesis). Holistically, this response includes clear and well-developed ideas throughout, typical of responses that receive scores of 3.

Test-taker Anchor Response 3 – Score: 3	Annotation
Each beaker should contain .5 liters of water. The beakers should be placed on a heat source and each should have its own thermometer. Each successive beaker should have a temperature of 5 degrees Farenheit above the previous beaker. The balance should be used to weigh out even distributions of salt. Each distribution of salt should weigh 1 gram. Data should be collected by straining the salt from the water and then the salt should be weighted to see how much was left undisolved. If more salt was left undisolved in the beakers with lower temperatures then the hypothesis would be valadated. If the amount of salt left did not vary across temperature or if more was left over in the higher temperature beakers then the hypothesis would be deemed invalid by this experiment. In either case the experiment would need to be repeated.	Score Explanation: This response contains a description of a logical experimental design, temperature is varied and equal amounts of salt are distributed to each beaker, (Each beaker should contain .5 liters of water…Each successive beaker should have a temperature of 5 degrees Farenheit above the previous beaker…Each distribution of salt should weigh 1 gram.). There is a well-formulated data collection method (. . . straining the salt from the water and then the salt should be weighed to see how much was left undisolved) which explains how to extract and measure the remaining salt from each beaker. The second paragraph includes a complete explanation of the criteria for evaluating the hypothesis because it connects the results back to the hypothesis by clearly describing what the results need to look like in order for the hypothesis to be validated (If more salt was left undisolved in the beakers with lower temperatures then the hypothesis would be valdated). Although this response lacks some of the details typical of responses that receive scores of 3, all of the ideas presented are clear and logical. Therefore, the response receives a score of 1.

Test-taker Anchor Response 4 – Score: 2	Annotation
1. Measure out three sets of 100 grams of salt using the balance. 2. Get three beakers and put 500 grams of water into each using th balance. 3. Get three thermoeteres and put on in each beaker. 4. Keep beaker one at room temp. 5. Heat beaker two to 70 degrees Celcius 6. Heat beaker three to 200 degrees Celcius 7. Add the 100 grams of salt to each of the beakers and gently stir around with the thermometer. 8. Observe how much salt is left in the beaker after 2 minutes of stirring the salt or if there is none left how quickly the salt vanished. 9. Record observations on a piece of paper When evaluating your hypothesis, you must first look at what you did. Which beaker has the most salt and which has the least? The salt didn't mysteriously vanished and in this expiriment there was no chemical reaction. The missing salt has dissolved into the water. When done correctly the beaker at room temp should have the most amount of visible salt left in the beaker. This means the solubility of salt in water increase as the temp of water increases.	Score Explanation: This response describes an experimental design that demonstrates use of a manipulated variable, temperature, (…put 500 grams of water into each…Keep beaker one at room temp…beaker two to 70 degrees…beaker three to 200 degrees…Add the 100 grams of salt to each of the beakers and gently stir.) Additionally, the test-taker includes a logical data collection method (Observe how much salt is left in the beaker). The criteria for evaluating the hypothesis is incomplete (. . . beaker at room temp should have the most amount of visible salt left in the beaker*)* since the results are not linked back to the hypothesis of the experiment. Although the incomplete criteria for evaluating the hypothesis lacks the details typical of responses that receive scores of 2, the strength of the experimental design clearly demonstrates the test-taker's ability to describe an experimental set-up to test a specific hypothesis. Therefore, it receives a score of 2.

Test-taker Anchor Response 5 – Score: 2	Annotation
First decide upon four increases water temperatures that you wish to test. Fill each of the four beakers with equal amounts of water (about 12oz). Then heat each beaker to the desired temperature, using the thermometer to maintain the stability of the temperature. Once the desired heat level is achieved in each beaker, add three tablespoons of salt to each one and stir for five seconds. Using your eyes and possibly a magnifying glass to observe, record your results. With this amount of salt it should be plain to see in which beaker it better dissolved. If it is not plain to see, drain the beakers using a very fine strainer and weigh the remaining salt on the balance and record your varying results. The beaker which had the least remaining salt, dissolved the best.	Score Explanation: This response describes a logical experimental design (Fill each of the four beakers with equal amounts of water…heat each beaker to the desired temperature…add three tablespoons of salt to each one….). The data collection method is well-reasoned because it describes how to collect the data from each beaker (. . . drain the beakers using a very fine strainer and weigh the remaining salt on the balance). Finally, the response has a logical, but not complete, criteria for evaluating the hypothesis (The beaker which had the least remaining salt, dissolved the best) since the temperature is not is not linked to the results or to the hypothesis. The logical data collection method corresponds with the experimental design, which helps compensate for the incomplete criteria for evaluating the hypothesis. Therefore, the response receives a score of 2.

Test-taker Anchor Response 6 – Score: 2	Annotation
Before beginning the experiment, the researcher must have salt, water, beakers , a heat source, and a thermometer. To set up, he must have a control and a variable. So, pour salt into water (in beaker#1) that is room temperature (check with the thermometer), and this will be the control for the experiment. To collect the data, the researcher will have three other beakers, with varying temperatures, beaker #2 being the coolest of the three, and beaker #4 being the hottest. After reaching the desirable temperatures by checking with the thermometer, the scientist can pour salt in each of the beakers, and then measure the solubility of the salt in each beaker. He can compare beakers 2, 3, and 4 to the control beaker, and see if there are any changes. To evaluate the hypothesis, the researcher can check each beaker, and see where the salt has dissolved most effectively. If his hypothesis is correct, the hottest beaker, beaker #4, will be the beaker where the salt dissolved most effectively.	Score Explanation: This response contains a description of a logical experimental design. Temperature is varied (…he must have a control and a variable…pour salt into water...that is room temperature,…three other beakers, with varying temperatures…); however, the poorly formulated data collection method (measure the solubility) does not describe **how** the solubility will be measured. Additionally, the response contains a weak explanation for the criteria for evaluating the hypothesis (If his hypothesis is correct, the hottest beaker, beaker #4, will be the beaker where the salt dissolved most effectively). The phrase "most effectively" is not quantified and therefore too imprecise in this response. Although the response demonstrates some weaknesses in formulating a logical data collection method and describing the criteria for evaluating the hypothesis, the quality of the experimental design is so strong in that it clearly demonstrates use of a manipulated variable. Therefore, this response receives a holistic score of 2.

Test-taker Anchor Response 7 – Score: 1	Annotation
Experimental Overview: This experiment will serve to test the hypothesis that as water temperature increases the solubility of salt in water also increases. Samples of salt added to water at set temperatures will be taken to provide data to help evaluate the stated hypothesis. Procedure: 1) Xgrams of salt will be weighed with a balance and set asided to later be added to the beakers of water. 2) Beakers will be filled with Xoz of water and placed on their heating sources. One beaker will remain unheated(room temperature) as a experimental control. 3) Each beaker will be heated to the experimental temperature. When the targeted temperature is reached and measured with the thermometer the salt will them be stirred in an the solubility will be measured. 4) Data from the experiment is to be recorded in a table to later be evaluated. Evaluation: Data collected will be compared to the control beaker. Any experimental values that stray from the measurement taken in the control beaker will determine wether or not a change in solubility occurred. For accurate results the experiment should be repeated multiple times under the same conditions.	Score Explanation: This response receives a score of 1 because while the response describes a logical experimental design demonstrating use of a manipulated variable, temperature (Xgrams of salt will be weighed…Beakers will be filled with Xoz of water...Each beaker will be heated to the experimental temperature…) there is a poorly formulated data collection method (solubility will be measured) . The data collection method does not describe **how** the solubility will be measured. The minimal explanation of the criteria for evaluating the hypothesis is also weak because the explanation is vague in terms of explaining how the individual results from the experiment will support or reject the hypothesis. (Data collected will be compared to the control beaker. Any experimental values that stray from the measurement taken in the control beaker will determine wether or not a change in solubility occurred*)*. While the well-developed experimental design compensates for the weak data collection method and minimal explanation of the criteria for evaluating the hypothesis, the ideas in this response are largely unclear and incomplete. Therefore, this response receives a score of 1.

Test-taker Anchor Response 8 – Score: 1	Annotation
Experimental setup: Set up three stations. Each station needs the same sized beaker filled with 150ml of water and 10 g of salt ready to be put in the water. Heat one beaker to 20 degrees Celsius, another to 40 degrees Celsius, and the third to 60 degrees Celsius. Each beaker should have a thermometer to monitor the temperature. Procedure: Once each beaker reaches the desired temperature, place the 10 grams of salt in the water, stir once, and wait one minute. I'm sorry I honestly don't know how to meaure solubility with only these materials present.	Score Explanation: This response describes a logical experimental design demonstrating use of a manipulated variable, temperature (…same sized beaker filled with 150ml of water and 10 g of salt ready…Heat one beaker to 20 degrees, another to 40 degrees…place the 10 grams of salt in the water…), However, the response does **not** attempt to describe a method for data collection or an explanation of the criteria for evaluating the hypothesis. Although the response completely lacks two of the components, this is compensated for by a logical experimental design that clearly demonstrates use of a manipulated variable. Therefore, this response earns a score of 1.

Test-taker Anchor Response 9 – Score: 1	Annotation
Title: The Effects of Temperature on Solubility Hypothesis: If the water temperature increases, then the solubility of the salt in the water also increases. Independent Variable: Solubility Dependent Variable: Temperature Procedure: 1. Gather 5 beakers and label them from A to E. 2. Each beaker will hold approximately half a liter of water within each. 3. Use appropriate equipment to heat/cool beakers when necessary. Beaker A will hold its water at a temperature of 0 degrees Celcius. Beaker B will hold its water at a temperature of 20 degrees Celcius. Beaker C will hold its water at a temperature of 40 degrees Celcius. Beaker d will hold its water at a temperature of 60 degrees Celcius. Beaker D will hold its water at a temperature of 80 degrees Celcius. Beaker E will hold its water at a temperature of 100 degrees Celcius. 4. Pour 10 grams of salt (measured out with the use of the balance) within each beaker and let sit for 5 minutes. 5. Measure solubility and record results.	Score Explanation This response contains a description of a logical experimental design, temperature is varied (Beaker A will hold its water at a temperature of 0 degrees Celcius. Beaker B will hold its water at a temperature of 20 degrees Celcius. Beaker C will hold its water at a temperature of 40 degrees Celcius.). The poorly formulated data collection method (Measure solubility…) does not describe **how** the solubility will be measured. Additionally, this response does **not** provide an explanation of the criteria for evaluating the hypothesis. Although the test-taker fails to address two components of the prompt, the logical experimental design compensates in that it clearly demonstrates use of a manipulated variable. Therefore, this response earns a score of 1.

Test-taker Anchor Response 10 – Score: 0	Annotation
Experimental setup: A beaker is filled halfway with water and an ounce of salt is added. A pan is filled with water and placed on the hot plate, heated to 300 degrees. Procddure for data collection methods: The beaker is placed in the hot water for five (5) minutes. At one minute intervals, the heat is increased by 50 degrees and the density of the salt is observed. Criteria for evaluating the hypothesis: The experiment proves that the hotter the temperature of the water, the faster the salt in the water dissolves or the salt is more soluble.	Score Explanation: This response does not describe an acceptable experimental design (A beaker is filled halfway with water and an ounce of salt is added .). There is an illogical data collection method (At one minute intervals, the heat is increased by 50 degrees and the density of the salt is observed.) and the response does not include an acceptable explanation for evaluating the hypothesis (. . . hotter the temperature of the water, the faster the salt in the water dissolves or the salt is more soluble.). Holistically, the response is very weak in that all of the test-takers ideas are illogical and scientifically inaccurate. Therefore, this response is too weak to earn a score.

Test-taker Anchor Response 11 – Score: 0	Annotation
Exp. Setup: find out which tempurature water disolves the salt. Proc. For data: put the salt inside two different cups of water one cold and one hot. Criteria for eval. Hypothesis: See which cup dissolved the salt, the cold water or the hot. That should tell u whether temperature has anything to do with dissovability.	Score Explanation: This response does not describe an experimental design (…find out which tempurature water disolves the salt.). There is an illogical data collection method (. . . put the salt inside two different cups of water one cold and one hot) and the explanation for evaluating the hypothesis does not clearly specify how to validate the hypothesis (See which cup dissolved the salt, the cold water or the hot). Holistically, the response is very weak in that most of the test-takers ideas are unclear or illogical. Therefore, the response is too weak to earn a score higher than zero.

Test-taker Anchor Response 12 – Score: 0	Annotation
Experimental setup: salt and water, various beakers, heat sources and temperature. Thermometer. Procedure for data: Test hypothesis of solubility of salt and walter in different temperature from low temperature to high temperature Collect sample and results from each material and form a conclusion.	Score Explanation: This response does not describe an experimental design, it re-lists the materials given in the prompt (Experimental setup: salt and water, various beakers, heat sources, and temperature. Thermometer). The illogical data collection method (Test hypothesis of solubility of salt and walter in different temperature from low temperature to high temperature) fails to address how to collect data from the experiment. Additionally, there is no explanation of the criteria for evaluating the hypothesis (Collect sample and results from each material and form a conclusion). The ideas in this response are vague and do not address the specific hypothesis provided in the prompt. Therefore, this response earns a score of 1.

2014 GED® Program Free Practice Test – Science – Item #6

The following pages present the textual stimulus and the prompt for the *Ophiocordyceps unilateralis* SA item from the 2014 GED® program free practice test (question 6).

Stimulus

Tropical rain forests contain diverse communities of organisms with many interesting relationships. One such relationship connects parasitic fungi and their insect hosts. A type of parasitic fungus, called *Ophiocordyceps unilateralis,* disperses spores onto the forest floor, but cannot successfully grow on the ground. The fungus requires specific conditions and must grow inside of a specific ant species, called the host, to reproduce. The ants, various species of carpenter ant, make nests in the trees.

O. unilateralis feeds on and grows inside the insect host, and within a few days the fungus affects the insect's brain, The insect exhibits unusual behaviors such as wandering away from the colony to where light and humidity favor fungal growth. Just before dying, the insect bites into and firmly attaches itself to a plant. Then, the fungus slowly grows outward from the dead insect's head, producing a pod of spores that eventually bursts open. The spores fall to the ground, restarting the life cycle of the fungus.

Though this relationship may sound gruesome, researchers note that these parasitic fungi may help maintain biodiversity in the tropical rain forest. Some parasitic fungi may be host-specific, meaning that a fungus species only infects a particular type of insect. Scientists have observed that if an insect population begins to grow, more fungal infections occur, and then the insect population levels off again. This relationship may prevent overpopulation of the habitat by anyone insect species.

Prompt

Deforestation, or clearing away trees, is occurring in tropical rain forests.

Explain how deforestation could disrupt the life cycle of *Ophiocordyceps unilateralis* in tropical rain forests. Include multiple pieces of evidence from the text to support your answer.

Type your response in the box. This task may require approximately 10 minutes to complete.

Ophiocordyceps unilateralis Science Short Answer Scoring Guide

Question Overview: In this SA question, test-takers are required to give an explanation about the effects of deforestation on the particular species of fungus described in the textual stimulus. Then, they must cite specific evidence from that text that supports their explanation. Their general understanding of ecosystems and life cycles may help them provide a more precise explanation of how deforestation can be disruptive. This question tests learners' skill at the complex task of using, producing, and justifying a text-based line of reasoning by incorporating elements from the text into the presentation of their own ideas.

Scoring Guide: Each response is scored on the basis of two key elements. Each bullet below describes the *quality* of these elements typical of each score point.

3-Point Response

Response contains

- a clear and well-developed explanation of how deforestation could disrupt the life cycle of *Ophiocordyceps unilateralis* In tropical rain forests
- complete support from the passage

2-Point Response

Response contains

- an adequate or partially articulated explanation *of* how deforestation could disrupt the life cycle of *Ophiocordyceps unilateralis* in tropical rain forests
- partial support from the passage

1-Point Response

Response contains

- a minimal or implied explanation of how deforestation could disrupt the life cycle of *Ophiocordyceps unilateralis* In tropical rain forests
- minimal or implied support from the passage

0-Point Response

Response Includes

- no explanation of how deforestation could disrupt the life cycle of *Ophiocordyceps unilateralis* in tropical rain forests
- no support from the passage

Non-scorable Responses (Score of 0/Condition Codes)

Response exclusively contains text copied from source text(s) or prompt
Response demonstrates that the that test-taker has read neither the prompt nor the source text(s)
Response is incomprehensible
Response is not in English
Response has not been attempted (blank)

Science Short Answer Responses and Annotations for *Ophiocordyceps unilateralis*

Text from the responses quoted within the annotations is highlighted in yellow in both the annotations and the test-taker response to help you quickly identify specific elements of each response that helped SMEs score them appropriately. However, keep in mind that each response must be considered as a whole, and these highlighted excerpts are notable mostly because they show specific examples of qualities common to responses.

Test-taker Anchor Response 1 – Score: 3	**Annotation**
Deforestation destroys the environment where thousands of species of animals flourish Including Ophiocordyceps. Ophiocordyceps rely heavily on the environment to survive for two major reasons. First of all, Ophiocordyceps often find hosts in Carpenter Ants which build their nests high up in the trees of rainforests. When deforestation occurs, Carpenter Ants lose their nests and homes which would likely result in the diminishment of their species. This would disrupt the Ophiocordyceps species significantly as Ophiocordyceps cannto survive without a host - without the Carpenter Ants, there would be no Ophiocordyceps. The other reason that Ophiocordyceps would suffer is because without the trees, there would be nothing for them to climb to reach greater amounts of light and less humidity. While lack of trees would lead to more light reaching the ground, the Issue of humidity affecting the Ophiocordyceps would still exist. With tall trees, the Ophiocordyceps are able to reach heights with less humidity hut deforestation would leave the Ophiocordyceps without a way to escape the humidity ultimately slowing the growth of the fungus. In conclusion, deforestation would have a very significant impact on the life cycle of the Ophiocordyceps for without trees there would be no hosts for the Ophiocordyceps to grow and without a way to escape humidity there would be a slowing of growth.	Score Explanation: This 3-polnt response explains how deforestation could disrupt the life cycle of *Ophiocordyceps unilateralis* by stating, ("Deforestation destroys the environment where thousands of species of animals flourish Including Ophiocordyceps.") This statement makes the connection between the destruction of the environment and its negative effect on the life cycle of Ophiocordyceps unilateralis. The explanation is supported with the following piece of evidence, ("Ophiocordyceps often find hosts in Carpenter Ants which build their nests high up in the trees of rainforests.") This piece of evidence links the trees with the living environment of the Ophiocordyceps. The explanation is further supported with a second piece of evidence, ("The other reason that Ophiocordyceps would suffer is because without the trees, there would be nothing for them to climb to reach greater amounts of light and less humidity.") This piece of evidence links the explanation

Test-taker Anchor Response 1 – Score: 3	Annotation
	(see comments on the previous page) of a loss of environment back to this statement in the passage which describes how the insects need the light and humidity because those conditions favor growth.

Test-taker Anchor Response 2 – Score: 2	Annotation
Ophiocordyceps unilateralis feed of the carpenter ant, which nests In the trees. Deforestation will cause many carpenter ant to die because of the lack of homes. As a result O. unilateralis lose many hosts to feed off of, and in turn reproduction is disrupted.	Score Explanation: This response explains how deforestation could disrupt the life cycle of Ophiocordyceps unilateralis by stating, ("Deforestation will cause many carpenter ant to die because of the lack of homes.") This statement describes how the destruction of the habitat or "home" of the Ophiocordyceps unilateralis has a negative effect on the life cycle of Ophiocordyceps. The explanation is supported with the following piece of evidence, "(...feed of the carpenter ant, which nests in the trees.)" This evidence, which is taken from the last sentence of the paragraph, provides an indirect reference as to how deforestation will destroy the "home" of the carpenter ant, which will in turn affect the Ophiocordyceps unilateralis. However, this response contains only partial support from the passage and therefore it receives a score of 2.

Test-taker Anchor Response 3 – Score: 1	Annotation
WITHOUT THE TRESS OPHIOCORDYCEPS UNILATERALIS CANNOT GROW BECAUSE THEY NEED THE TREES TO DISPERSE SPORES ONTO THE FOREST FLOOR IN ORDER TO GROW AND REPRODUCE.	Score Explanation: This response gives an explanation of how deforestation could disrupt the lifecycle of the Ophiocordyceps unilateralis by stating, ("BECAUSE THEY NEED THE TREES TO DISPERSE SPORES ONTO THE FOREST FLOOR IN ORDER TO GROW AND REPRODUCE.") The response explains how the Ophiocordyceps unilateralis requires the trees in order to continue its lifecycle. However, it does not include any supporting textual evidence from the passage; therefore, this response receives a score of 1.

2014 GED® Program Free Practice Test – Science – Item #12

The following pages present the stimulus and the prompt for the *Farmer's Hypothesis Experimental Design* SA question from the 2014 GED® program free practice test (question 12).

Stimulus Material

A farmer purchased 30 acres of farmland. The farmer calculated that the average topsoil thickness on the farmland is about 20 centimeters.

The farmer wants to maintain the thickness of the soil on this farmland by reducing erosion. The farmer plans to test the effectiveness of two different farming methods for reducing soil erosion.

Method 1: No-till (planting crops without plowing the soil)
Method 2: Winter cover crop (growing plants during the winter that are plowed into the soil in spring)

The farmer hypothesizes that using either method will reduce erosion compared to using traditional farming methods (plowing and no cover crop).

Prompt

Design a controlled experiment that the farmer can use to test this hypothesis. Include descriptions of data collection and how the farmer will determine whether his hypothesis is correct.

Type your response in the box. This task may require approximately 10 minutes to complete.

Farmer's Hypothesis Experimental Design Science Short Answer Scoring Guide

Question Overview: The purpose of this SA question is to provide the test-taker with an opportunity to design a scientific investigation. The test-taker must understand and apply the fundamentals of scientific investigation design in his or her response. In addition, the test-taker must describe an experimental set-up for a controlled experiment, which includes a description of the methods used to collect data. The test-taker must also justify his or her line of reasoning used to determine whether the hypothesis is valid. Correctly completing the multi-step process required by this question involves several critical thinking skills, including identifying a research question, designing a scientific investigation, and justifying a line of reasoning.

Scoring Guide: Each response is scored on the basis of three key elements. Each bullet below describes the *quality* of these elements typical of each score point.

3- point Response

Response contains

- a well-formulated, complete controlled experimental design
- a well-formulated data collection method
- a well-formulated, complete explanation of the criteria for evaluating the hypothesis

2-Point Response

Response contains

- a logical controlled experimental design
- a logical data collection method
- a logical explanation of the criteria for evaluating the hypothesis

1-Point Response

Response contains

- a minimal experimental design
- a minimal or poorly-formulated data collection method
- a minimal or poorly-formulated explanation of the criteria for evaluating the hypothesis

0-Point Response

Response contains

- an illogical or no experimental design
- an illogical or no data collection method
- an illogical or no explanation of the criteria for evaluating the hypothesis

Non-scorable Responses (Score of 0/Condition Codes)

Response exclusively contains text copied from source text(s) or prompt
Response demonstrates that the that test-taker has read neither the prompt nor the source text(s)
Response is incomprehensible
Response is not in English
Response has not been attempted (blank)

Science Short Answer Responses and Annotations for *Farmer's Hypothesis Experimental Design*

Text from the responses quoted within the annotations is highlighted in yellow in both the annotations and the test-taker response to help you quickly identify specific elements of each response that helped SMEs score them appropriately. However, keep in mind that each response must be considered as a whole, and these highlighted excerpts are notable mostly because they show specific examples of qualities common to responses.

Test-taker Anchor Response 1 – Score: 3	**Annotation**
The farmer would have to set up 3 experiments. The first would be a years worth of traditional farming methods (plowing and no cover crop) on 5 x 5 acres of land. He would have to measure the top soil in every month throughout the year and record it in a data table. For the second experiment the farmer would have to farm a plot of land 5x5 acres using a no-till plan. He would have to measure the top soil every month for a year and record it in a data table. Finally the farmer would farm a 5x5 acres of land with winter cover crop and measure the top soil every month and record it in a lab table. At the end of the year the farmer would have to compare the 2 methos agaisnt the traditional methid and determine ifhe is correct	Score Explanation: This response earns all three points because it includes a complete description of the experiment and includes the controlled variable, ("The farmer would have to set up 3 experiments. The first would be a years worth of traditional farming methods (plowing and no cover crop) on 5 x 5 acres of land.") The response also describes data collection methods for the control group and experimental group by stating that the farmer ("... He would have to measure the top soil every month for a year and record it in a data table.") Finally, the response provides an explanation of how the farmer will determine if his hypothesis is correct, ("At the end of the year the farmer would have to compare the 2 methos agaisnt the traditional methid and determine ifhe is correct")

Test-taker Anchor Response 2 – Score: 2	Annotation
The farmer could separate the land into two sections (15 acres each), and use one method on each section over a two season period. Over the two season period he would record how much soil was left after using each method, comparing the results to each other and the traditional farming method.	Score Explanation: This response include a logical description of the experiment, ("The farmer could separate the land into two sections (15 acres each), and use one method on each section…") While the response does not include the controlled variable in this initial description of the experiment, in the last sentence of the response the writer states, ("...comparing the results to each other and the traditional farming method.") This statement demonstrates that the writer understands the connection and importance of having the controlled variable as part of the experiment and data collection methods. The response also describes a logical data collection method by stating, ("Over the two season period he would record how much soil was left after using each method…") However, this response only provides an implied, logical explanation of the criteria for evaluating if the hypothesis is correct by stating, ("...comparing the results to each other and the traditional farming method.") While it is clear that the response is attempting to connect the results of the experiment with an evaluation of the hypothesis, this is not a complete statement.

Test-taker Anchor Response 3 – Score: 1	Annotation
To test his hypothesis the farmer should divide his land into three equal parts one for the first method, one for the second method and one for the controle group. In the first part he divided he should test method one and keep a record of the process and the results. In the second part he divided he should test the second method and keep a record of the process and the results. In the third part that he divided he should have the controle group where he would use the traditional method keep a record of tre process and the results,then compare the records he has collected identify the different results, make an annalasys and decide which method is the best way to prevent soil erosion.	Score Explanation: This response includes a description of the experiment, ("...the farmer should divide his land into three equal parts one for the first method, one for the second method and one for the controle group.") The response also describes a poorly formulated data collection method by stating, ("In the first part he divided he should test method one and keep a record of the process and the results.In the second part he divided he should test the second method and keep a record of the process and the results. In the third part that he divided he should have the controle group where he would use the traditional method keep a record of tre process and the results …") While the response is describing the collection of data in all three sections of land, the phrase "keep a record of tre process and results" is NOT a clear statement describing what type of data will be collected. The response also describes a minimal explanation of the criteria for evaluating the hypothesis by stating ("...compare the records he has collected identify the different results, make an annalasys and decide which method…".)

Automated Scoring of Constructed Response Items on the 2014 GED® Test

The 2014 GED® test contains four Constructed Response (CR) items: one 45-minute Extended Response (ER) item on the Reasoning Through Language Arts (RLA) module, one 25-minute Extended Response (ER) item on the Social Studies module, and two 10-minute Short Answer (SA) items on the Science module.

Logistically, the ER item in RLA is in its own separately-timed section of the test at the end of the first half of the RLA module (prior to a 10-minute break). The ER item in the Social Studies test is in its own separately-timed section that appears as the last item of the Social Studies module. The Science SA items are distributed within the 90-minute Science module and are not timed separately—test-takers use their time-management skills to monitor their use of time on those items and are given guidelines as to approximately how much writing is expected in those responses (the test-taker is instructed to take up to about 10 minutes to read the question, and formulate, write, and edit their answer).

It was a critical goal of GED Testing Service to incorporate CR items into the design of the 2014 GED® test because these types of items are a key method of assessing a test-taker's higher order thinking skills as well as their skills in expressing themselves clearly in their own words. To ensure that the results of testing are available to test-takers in the quickest timeframe possible (because adults usually do not have the luxury of waiting days or weeks for their test results to be finalized), GED Testing Service will be scoring CR items using an automated scoring engine, supplemented by human scorers as necessary, described in more detail below.

Great strides have been made in automated scoring over the last decade, and the use of automated scoring is intended to replicate the human scoring process. However, the automated scoring engine will need to be supplemented by human scorers in certain circumstances. Automated scoring is not fully developed enough to result in reliable scoring in the area of mathematics, so GED Testing Service elected not to incorporate CR items into the Mathematical Reasoning test. We hope in the future to be able to build those item types into the test as the technology develops and matures in the future.

The following description applies equally to all CR items, whether ER or SA.

During the item development process, experts in automated scoring are involved from the outset, rather than being brought into the process after items have already been authored. This collaborative consultation and review helps ensure that responses have a high likelihood of being reliably scored by the automated engine. For example, questions that do not provide adequate instruction to the test-takers about what information they should include in their answers sometimes produce a wide and/or unpredictable range of responses that both people and computers can have difficulty in scoring consistently. Creating item stems that focus the test-taker on the specific expectations of the item is important so that the item can both validly assess the intended content specification and also have a high probability of being scored appropriately and reliably both by humans and computer.

Once items have been written, reviewed by both scoring and content experts, and finalized, they are field-tested. In the case of the initial forms for the 2014 GED® test, thousands of test-takers in locations across the U.S. in the summer and fall of 2012 participated in the field-

testing. The test-takers that were recruited to participate matched the profile of our adult GED® test-taking population. At the conclusion of field testing, the written responses to the CR items were examined and a sample of test-taker responses was selected for each of the items. Teams of content experts reviewed the responses in a process known as "rangefinding." The purpose of rangefinding is to determine range and variety of responses that fulfill each score point as defined on the rubric that is very carefully constructed and designed to guide the overall evaluation of responses. This standard best-practice procedure for the scoring of CR items results in the selection of exemplar responses at each score point. These responses are used to build anchor sets (human scorers' official guide that is used in evaluating test-taker responses), practice sets (sets of responses used in training human scorers), and qualification sets (sets of responses that scorers take in a "quiz" in which they must match their scores to "true scores" given during rangefinding to qualify to appropriately and reliably score CR items).

When these materials have been compiled and scorer training is complete, all of the test-taker responses from the field test are scored by humans, using the "double read with resolution" approach. This scoring model entails each and every response being read and scored independently by no fewer than two individuals. If the scores applied by the two different scorers are in exact agreement, the score for that response is final. If the two scores differ by only a single point, they are averaged and rounded up, effectively resulting in acceptance of the higher score point. If the scores differ by more that one point ("non-adjacent scores"), the response is read by a scoring leader (an expert scorer) who determines the correct score for that response in a process called "resolution." Because the ER items are scored across three key traits, each of which contains multiple dimensions that are considered together in a compensatory manner (meaning that a response that is particularly strong in one dimension can still receive a higher score even if it is weaker in other dimensions), each ER response is actually read by no fewer than six people. That is, each scorer is trained to score only one rubric trait, and two scorers trained on each of the three traits read each response. Therefore, it is possible for a single ER response to be read by up to nine people, if the first two scores on all three traits are non-adjacent. This process ensures that the human scoring process produces the highest quality results and data.

When the scoring of all of the responses generated through field-testing is complete, a team of content experts, psychometricians and automated scoring experts reviews the range of scores for each constructed response item. At that time, some items are rejected because they do not meet the minimum criteria for inclusion on any operational 2014 GED® test or GED Ready™: The Official Practice Test. Items that survive this process then are passed along to the scoring organization to train the automated scoring engine. Several hundred scored responses for each item are fed into the automated scoring engine. Then, several hundred more scored responses are used to test the reliability of scores generated by the automated engine. The engine evaluates each response on over 100 different dimensions in relation to the score that the response was given. Through this training and testing procedure, the automated engine "learns" how to score the items and is then able to replicate the scoring that was done by humans. Once this process is complete, data from the replication process is reviewed, and occasionally, if the scoring is determined to be insufficiently reliable to be used on an operational 2014 GED® test during this data review, some items may be allocated for use on GED Ready™ because the CR items on the practice test are always scored by humans.

Only CR items that successfully survive the entirety of this process are placed on operational 2014 GED® test forms. When the test goes live in 2014, test-takers will respond to the CR

items and their responses will be fed into the automated engine for scoring immediately upon completion of each individual content area test. Of course, there will be a slight delay in submission of responses for scoring in some testing situations, such as with tests administered within the corrections system, in which the testing center is Internet independent. In these situations, additional steps need to be taken to upload the raw testing data (e.g., the test-takers' responses themselves) via a secure Internet connection.

Based on the experience of GED Testing Service with automated scoring during the field testing and other test development processes, we expect the vast majority of test-taker responses (most likely 95 percent or greater) to be reliably scored by the automated scoring engine—in a process that is completed in nanoseconds. However, as with any process that involves the variability present in people's writing, there will be responses that the automated scoring engine will recognize as not fitting any type of response that was previously seen in the training of the engine. For example, an extremely short response that uses a great deal of advanced vocabulary might be unusual and therefore would be automatically flagged by the automated scoring engine as an "outlier" in need of human intervention for scoring. These outlier responses are securely routed electronically to a network of human scorers who have been trained to score the item using the anchor items and training sets created during the rangefinding process, as well as the scoring rubric that is used to provide overall guidance to the scoring process. These human scorers score the test-taker response using the "double read with resolution" framework that was also used to score the field test responses.

Although the human scoring process is efficient, it does require additional time. GED Testing Service is committed to returning test results and a score report to test-takers within three hours of the completion of each test. Of course, the vast majority of results would actually be ready immediately because of advantage of the speed of the automated scoring, but, in order to manage test-taker expectations and avoid situations in which one test-taker at a site receives a score immediately while another test-taker does not, a three-hour delay has been built into the process of delivering test scores.

Three additional quality control procedures have also been built into the automated scoring system to ensure that test-takers receive reliable and valid scores from this process.

First, when the test goes live in 2014, the program will implement a process known as the "Initial Analysis Period" (IAP). The purpose of the IAP is to provide final validation of the automated scoring engine and its performance with the adult population of GED® test-takers. During the IAP, all CR responses will be scored both by the automated scoring engine and by human scorers (using the "double-read with resolution" model as appropriate). This ensures that all test-takers are being evaluated fairly and that the automated scoring engine is operating properly.

Second, an audit procedure will be conducted on an on-going basis, in which a percentage of all test-taker responses scored by the automated engine will be reviewed by human scorers. This audit will be *in addition* to the scoring of "outliers" described above, and will help to ensure the ongoing accuracy of the system.

Third, an automatic rescore process is being implemented. This process flags tests that have a failing score within a predetermined margin, such that if the CR scores on the test would have been higher, the final score result would have changed from "fail" to "pass." The CR responses

on these flagged tests will also be automatically routed to human scorers for evaluation so that the results from the automated scoring engine can be confirmed or adjustments made.

Because of the extreme care that GED Testing Service is taking with implementation of the automated scoring engine, in combination with human scoring and audit procedures, we are highly confident that our approach will produce high quality results with reliable and valid test scores for our test-takers. Due to ongoing involvement of human scorers in the scoring process (through the IAP, evaluation of outlier responses, the audit procedure, and the automatic rescore), the database of known response types will grow over time. This expanded response base will be used to periodically retrain the automated scoring engine to further improve its performance.

Finally, another key benefit of using the automated scoring engine technology is that it allows GED Testing Service to integrate specific feedback on test-takers' performance on the extended response and SA items right into the standard score report—a useful new feature that has never been possible in the past with the paper-based scoring system. This valuable process is part of GED Testing Service's effort to create a more learner-based testing system that will help guide test-takers to continuously improve their performance.