Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $S: \mathbb{R}^m \rightarrow \mathbb{R}^p$ be linear transformations. Then the composition of functions $S \circ T$ is a function from $\mathbb{R}^n$ to $\mathbb{R}^p$ defined by $(S \circ T)(\vec{x}) = S(T(\vec{x}))$ for all $\vec{x} \in \mathbb{R}^n$.

Note that $T$ is applied to a vector in $\vec{x} \in \mathbb{R}^n$. Then $T(\vec{x})$ is in $\mathbb{R}^m$. We then apply $S$ to $T(\vec{x})$. Then $S(T(\vec{x}))$ is in $\mathbb{R}^p$. So $S \circ T$ is applied to a vector $\vec{x}$ in $\mathbb{R}^n$ and $(S \circ T)(\vec{x})$ ends up in $\mathbb{R}^p$; thus we may write $S \circ T: \mathbb{R}^n \rightarrow \mathbb{R}^p$.

**Theorem 2.10.** Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a linear transformation with $m \times n$ matrix representation $A$, and let $S: \mathbb{R}^m \rightarrow \mathbb{R}^p$ be a linear transformation with $p \times m$ matrix representation $B$. Then $S \circ T$ is a linear transformation with matrix representation $BA$.

**Proof.** To show that $S \circ T$ is linear, let $\vec{x}, \vec{y} \in \mathbb{R}^n$ and let $c$ be a scalar. Then because $T$ and $S$ are linear, we have

$$(S \circ T)(c \vec{x} + \vec{y}) = S(T(c \vec{x} + \vec{y})) = S(cT(\vec{x}) + T(\vec{y}))$$

$$= cS(T(\vec{x})) + S(T(\vec{y}))$$

$$= c(S \circ T)(\vec{x}) + (S \circ T)(\vec{y}).$$

Hence, $S \circ T$ is linear.

Next, $T(\vec{x}) = AX$ and $S(\vec{w}) = BW$, where $X$ is the column matrix form of $\vec{x}$ and $W$ is the column matrix form of $\vec{w}$. Then, $(S \circ T)(\vec{x}) = S(T(\vec{x})) = B \times (AX) = (BA)X$. So $BA$ is the matrix representation of $S \circ T$.

**Note:** Applying the composition to $X$ in matrix form as $(BA)X = B(AX)$ means that we apply $A$ first and then apply $B$, which is the same as applying $T$ first and then applying $S$. 
Example. Let \( T: \mathbb{R}^3 \to \mathbb{R}^2 \) and \( S: \mathbb{R}^2 \to \mathbb{R}^4 \) be defined by
\[
T(\langle x_1, x_2, x_3 \rangle) = (x_1 - x_2 + 2x_3, 2x_1 + x_2 - 4x_3)
\]
and
\[
S(\langle x_1, x_2 \rangle) = (2x_1 + 3x_2, 5x_1, 4x_2, 3x_1 - x_2)
\]
Give the matrix representation and function form of \( S \circ T \).

Solution. First, \( T: \mathbb{R}^3 \to \mathbb{R}^2 \) has the \( 2 \times 3 \) matrix representation \( A = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 1 & -4 \end{pmatrix} \) and \( S: \mathbb{R}^2 \to \mathbb{R}^4 \) has the \( 4 \times 2 \) matrix representation \( B = \begin{pmatrix} 2 & 3 \\ 5 & 0 \\ 0 & 4 \\ 3 & -1 \end{pmatrix} \). Thus, \( BA \) is defined as a \( 4 \times 3 \) matrix and
\[
BA = \begin{pmatrix} 8 & 1 & -8 \\ 5 & -5 & 10 \\ 8 & 4 & -16 \\ 1 & -4 & 10 \end{pmatrix}
\]
\( S \circ T: \mathbb{R}^3 \to \mathbb{R}^4 \), which in function form is
\[
(S \circ T)(\langle x_1, x_2, x_3 \rangle) = (8x_1 + x_2 - 8x_3, 5x_1 - 5x_2 + 10x_3, 8x_1 + 4x_2 - 16x_3, x_1 - 4x_2 + 10x_3)
\]

Theorem 2.11. Let \( T: \mathbb{R}^n \to \mathbb{R}^m \) and \( S: \mathbb{R}^m \to \mathbb{R}^p \) be linear transformations. If \( T \) and \( S \) are both one-to-one, then \( S \circ T \) is one-to-one.

Proof. Suppose \((S \circ T)(\bar{x}) = (S \circ T)(\bar{y})\) for some \( \bar{x}, \bar{y} \in \mathbb{R}^n \). We must show that \( \bar{x} = \bar{y} \). But we have \( S(T(\bar{x})) = S(T(\bar{y})) \). So because \( S \) is one-to-one, we must have \( T(\bar{x}) = T(\bar{y}) \). Then because \( T \) is one-to-one, we must have \( \bar{x} = \bar{y} \).

Theorem 2.12. Let \( T: \mathbb{R}^n \to \mathbb{R}^m \) and \( S: \mathbb{R}^m \to \mathbb{R}^p \) be linear transformations. If \( T \) and \( S \) are both onto, then \( S \circ T \) is onto.

Proof. First, \( S \circ T: \mathbb{R}^n \to \mathbb{R}^p \) has co-domain \( \mathbb{R}^p \). To show \( S \circ T \) is onto, we must show that every vector in \( \mathbb{R}^p \) has a pre-image in \( \mathbb{R}^n \). So let \( \bar{z} \in \mathbb{R}^p \). Because \( S \) is onto, there exists \( \bar{y} \in \mathbb{R}^m \) such that \( S(\bar{y}) = \bar{z} \). Then because \( T \) is onto, there exists \( \bar{x} \in \mathbb{R}^n \) such that \( T(\bar{x}) = \bar{y} \). Then \((S \circ T)(\bar{x}) = S(T(\bar{x})) = S(\bar{y}) = \bar{z}\). So \( \bar{x} \) is the pre-image of \( \bar{z} \) and thus \( S \circ T \) is onto.
Theorem 2.13. Let $T: \mathbb{R}^n \to \mathbb{R}^n$ and $S: \mathbb{R}^n \to \mathbb{R}^n$ be linear transformations, both of which are one-to-one and onto, having matrix representations $A$ and $B$, respectively. Then $S \circ T$ is also one-to-one and onto and $(S \circ T)^{-1} = T^{-1} \circ S^{-1}$. Moreover, the matrix representation of $(S \circ T)^{-1}$ is $(BA)^{-1} = A^{-1}B^{-1}$.

Proof. By Theorems 2 and 3, $S \circ T: \mathbb{R}^n \to \mathbb{R}^n$ is both one-to-one and onto and by Theorem 1 has matrix representation $BA$. Therefore, $S \circ T$ is invertible and the matrix representation of its inverse is $(BA)^{-1} = A^{-1}B^{-1}$, which is the unique representation of the composition $T^{-1} \circ S^{-1}$. Thus, $T^{-1} \circ S^{-1} = (S \circ T)^{-1}$.

So for $T: \mathbb{R}^n \to \mathbb{R}^n$ and $S: \mathbb{R}^n \to \mathbb{R}^n$, both one-to-one and onto, then $S \circ T: \mathbb{R}^n \to \mathbb{R}^n$ is also one-to-one and onto and thus has an inverse which is $T^{-1} \circ S^{-1}$. In other words, for $S \circ T$, we apply $T$ first and then apply $S$. To undo this process, we undo $S$ first and then undo $T$, which is precisely the action of $T^{-1} \circ S^{-1}$:

$$\tilde{z} = (S \circ T)(\tilde{x}) = S(T(\tilde{x})) \quad \text{(Apply } T \text{ first and then apply } S)$$

Now undo:

$$\tilde{z} = S(T(\tilde{x})) \quad \rightarrow \quad S^{-1}(\tilde{z}) = T(\tilde{x}) \quad \rightarrow \quad T^{-1}(S^{-1}(\tilde{z})) = \tilde{x}$$

(undo $S$ first and then undo $T$)

So if $\tilde{z} = (S \circ T)(\tilde{x})$,

then $\tilde{x} = (T^{-1} \circ S^{-1})(\tilde{z})$. 